

# Grooves & Gains: Music You Should Listen to When Making Financial Decisions

Rat Pack

2023-04-27

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Packages Required . . . . .	2
<b>3</b>	<b>Data Cleaning and Variable Creation</b>	<b>2</b>
<b>4</b>	<b>Descriptive Plots and Summary Information</b>	<b>3</b>
4.1	Boxplots . . . . .	3
4.2	Scatterplots . . . . .	7
4.3	T-Tests . . . . .	9
4.4	Correlations Plot . . . . .	10
4.5	Correlation Chart . . . . .	12
4.6	Bootstrap Test . . . . .	12
<b>5</b>	<b>Permutation Test</b>	<b>13</b>
<b>6</b>	<b>Multiple Regression</b>	<b>14</b>
6.1	Description of Plan . . . . .	14
6.2	Residual Plots . . . . .	14
6.3	Results and Discussion . . . . .	15
6.4	Results and Discussion . . . . .	16
6.5	Final Model . . . . .	19
6.6	Ancova . . . . .	19
<b>7</b>	<b>Conclusion/Summary</b>	<b>19</b>
<b>8</b>	<b>Works Cited</b>	<b>20</b>

## 1 Introduction

Research has shown that the presence of music can significantly influence our cognitive abilities as we perform our daily routines. For instance, some studies suggest that the tempo of music can impact the pace of customer responses, while calming melodies may enhance our mental capacities (Cockerton et al., 1997). Music choice, however, also has the ability to overwhelm our cognitive space, resulting in reduced efficiency (Thompson et al., 2012). This is observed most when both the task and the music demand a similar type of cognitive processing.

Part of our routines also involve making everyday financial decisions whether its buying groceries or buying stocks. These decisions involve different aspects of uncertainty which can either lead to either fruitful gains or unfortunate losses.

In this report, we analyze the factors affecting risk-taking in financial decisions and explore how music-related factors affect our economic decisions and improve decision-making to increase investment success.

## 2 Data

In this report, we utilized a behavioral questionnaire dataset from a study conducted on individuals related to both music and financial risk taking. The variables of relevance include:

- **Group:** Categorical. The type of music played for the subject. 0= Slow tempo, 1=Fast tempo, 2= No music.
- **Mood before:** Categorical. Subjects chose between 10 options that best represent their mood prior to study.
- **Lot1-3:** Continuous. Numerical value representing the subjects performance on the lottery task of the study.
- **Inv A-C:** Continuous. Numerical value representing the subjects performance on the portfolio-diversification task of the study.
- **Age:** Continuous. Subjects age in years.
- **Gender:** Categorical. Subjects gender. 0=Male, 1=Female.
- **Marital:** Categorical. Subjects marital status. 1=single, 2=Married, 3= Divorced, 4=Widowed, 5= Separated.
- **Kids\_no:** Continuous. Subjects number of children.
- **Birth\_order:** Continuous. Position in family.
- **Residence:** Place of residence.
- **Religion:** Religion status.
- **Overdradt:** Often current bank account overdrawn.
- **Mood after:** Categorical. Subjects chose between 10 options that best represent their mood after studying.
- **music\_rec:** Categorical. Is the music playing in the background familiar to the subject?
- **music\_like:** Categorical. Did the subject like the music?
- **music\_help:** Categorical. Does the subject think the music helped or disturbed them while they filled out the questionnaire?
- **music\_effect:** Categorical. How did the music affect the subjects concentration?
- **Earning:** Categorical. Does the subject have hearing problems?

### 2.1 Packages Required

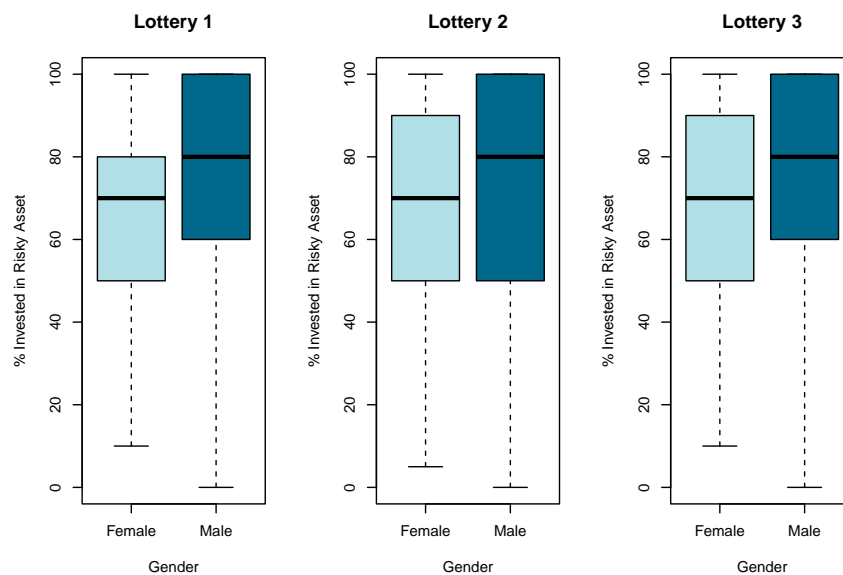
- `carData`
- `timechange`
- `libridate`
- `olsrr`
- `ggplot2`

## 3 Data Cleaning and Variable Creation

This R code performs data **cleaning** and **transformation** operations on a dataset named “data”. Some variables such as **Lot1** had NA values instead of 0s. To fix this, we ran code to identify NA values in the variable columns and replaced NA with ‘0’. Some of the categorical variables such as **Gender** used binary (0 or 1) rather than text. To fix this, we used `as.numeric()` function to change numbers to numerical format. Some categorical variables like **Gender** used binary (0 or 1) inputs instead of text. To fix this, we created new variables such as **invs\_return**, calculating expected return of participants’ investments using `rowSums()` and `rowMeans()`. Some variables such as **InvA** (“coins” invested in security A) do not reveal much about investment habits. The **Age** variable was not in numerical format which we corrected. Lastly, we ran code that identified and replaced misspelled variable names with corrected names.

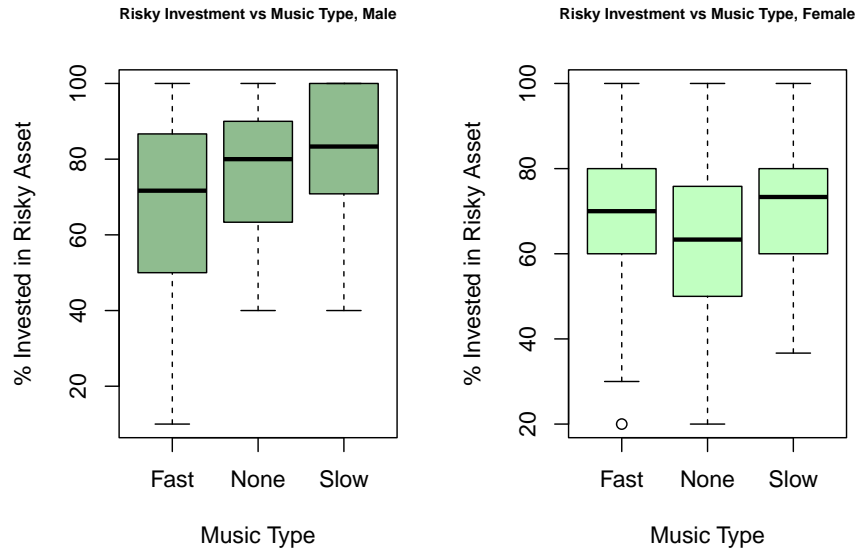
## 4 Descriptive Plots and Summary Information

### 4.1 Boxplots

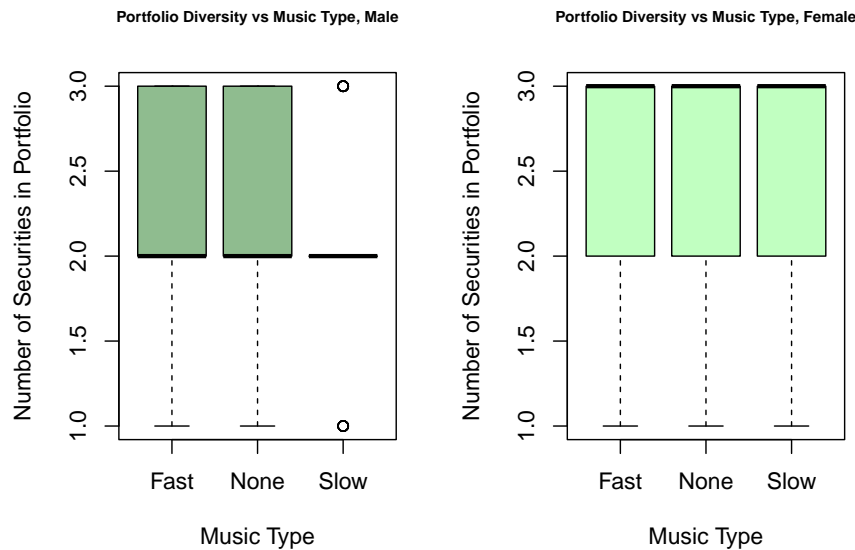


```
##
##  Welch Two Sample t-test
##
## data:  data$Lot_avg by data$gender
## t = -4.2324, df = 346.21, p-value = 2.965e-05
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
## 95 percent confidence interval:
##  -12.227829  -4.468774
## sample estimates:
## mean in group Female    mean in group Male
##           67.70238           76.05068
```

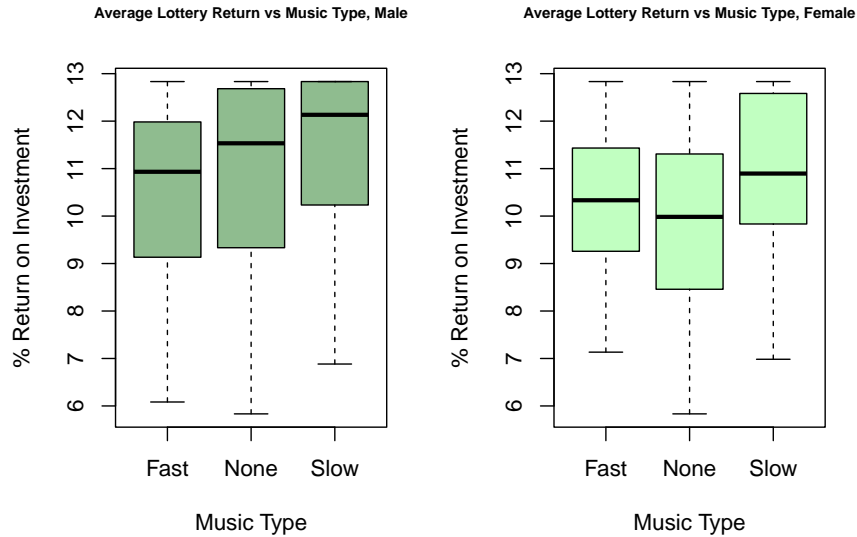
Based on these boxplots we observe that in Lottery 1, 2 and 3, Females tend to invest a lower percentage (around 40%) in the risky asset compared to males (around 80%). The pattern is consistent across all three lotteries, indicating that males tend to invest a higher percentage in the risky asset compared to females. Additionally, the p-value is extremely small (0.000029), which suggests that the difference in means between the Female and Male groups is statistically significant. The 95% confidence interval for the difference in means is (-12.23, -4.47), which does not include 0. This further confirms that the means are significantly different between the two groups. The sample estimates show that the mean “lot\_avg” for the Female group is 67.70, while for the Male group, it is 76.05. This aligns with the boxplot observations, where males tend to invest a higher percentage in the risky asset compared to females. Overall, both the boxplots and the t-test results indicate a significant difference in the investment behavior between males and females, with males investing a higher percentage in the risky asset compared to females across the three lotteries.



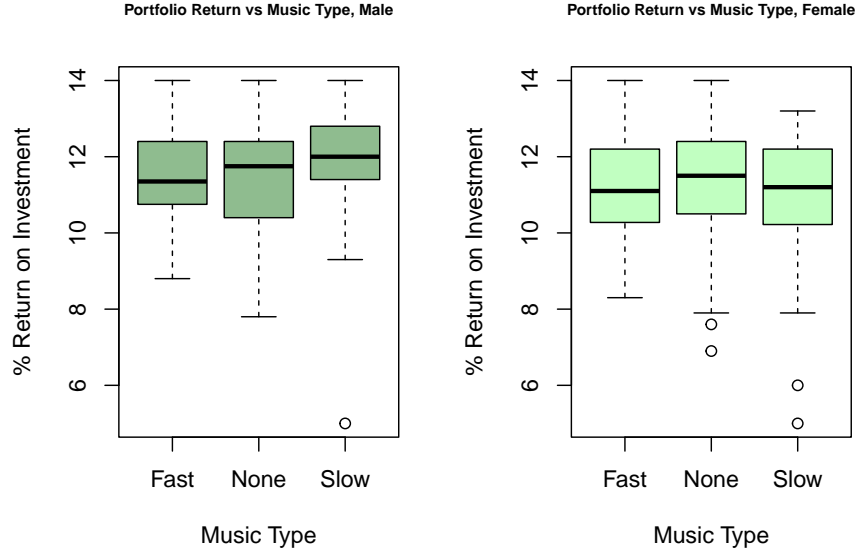
The boxplots display the relationship between music type (Fast, None, Slow) and the percentage invested in a risky asset, separately for males and females. For males: those who listened to slow music tended to invest the highest percentage (around 80-90%) in the risky asset, those who listened to no music invested a slightly lower percentage (around 60-80%) compared to slow music, and those who listened to fast music invested the lowest percentage (around 40-60%) in the risky asset. For females, the overall percentage invested in the risky asset is lower compared to males across all music types, similar to males those who listened to slow music tended to invest the highest percentage (around 60-80%) in the risky asset, and those who listened to no music or fast music invested lower percentages, with considerable overlap between the two groups (around 40-60%). Overall, for both genders, listening to slow music was associated with higher risk-taking behavior in terms of investment allocation compared to fast music or no music conditions. Additionally, females generally exhibited lower risk-taking behavior across all music types compared to males. These findings suggest that the type of music an individual listens to may influence their risk preferences and investment decisions, with slow music potentially promoting higher risk-taking tendencies, especially among males.



Based on the boxplots, for males, we observe that those who listened to fast music or no music had similar levels of portfolio diversity, with the median around 2.0-2.5 securities in their portfolio, while those who listened to slow music appeared to have slightly higher portfolio diversity, with the median around 3.0 securities. However, there is an outlier in the slow music group with a very high number of securities (around 6.0). For females, the distribution of portfolio diversity is very similar across all three music types (fast, none, slow), with the median portfolio diversity around 2.0-2.5 securities for all three groups, and little variation or difference is observed based on the music type for females. Overall, for males, listening to slow music seems to be associated with slightly higher portfolio diversity compared to fast music or no music conditions. However, this effect is not evident for females, where portfolio diversity appears to be unrelated to the type of music.

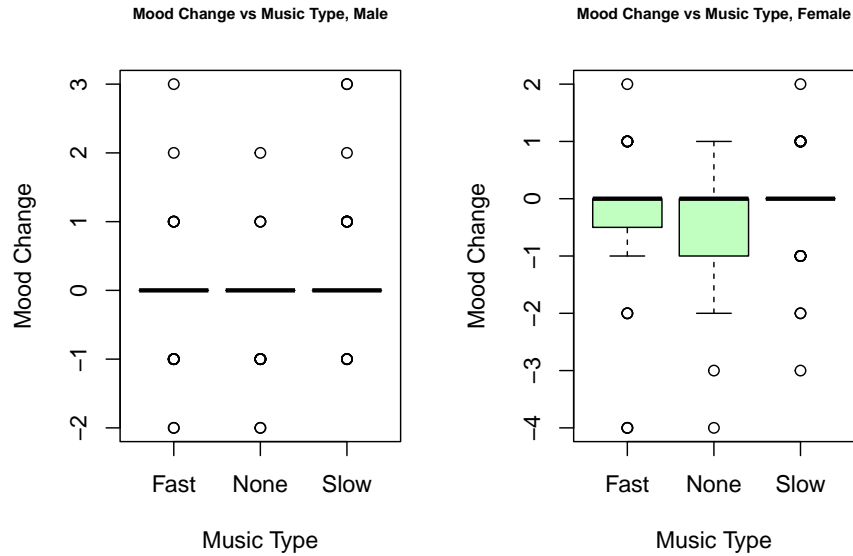


Based on the boxplots illustrating the average lottery return across different music types for males and females, we observe the following for males: those who listened to slow music tended to have the highest average lottery returns, with the median around 12-13%; those who listened to no music had slightly lower average returns compared to slow music, with the median around 10-11%; and those who listened to fast music had the lowest average lottery returns, with the median around 8-9%. For females, the distribution of average lottery returns is more compressed compared to males across all music types. Those who listened to slow music also tended to have the highest average returns, with the median around 10-11%. There is considerable overlap in the distributions for those who listened to fast music or no music, with the medians around 9-10%. Overall, for both males and females, listening to slow music was associated with higher average lottery returns compared to fast music or no music conditions. This trend is more pronounced for males, where the differences in median returns across music types are larger. These findings align with the previous observations on risky asset allocation, where slow music was linked to higher risk-taking behavior, potentially leading to higher returns on average from the lottery investment.

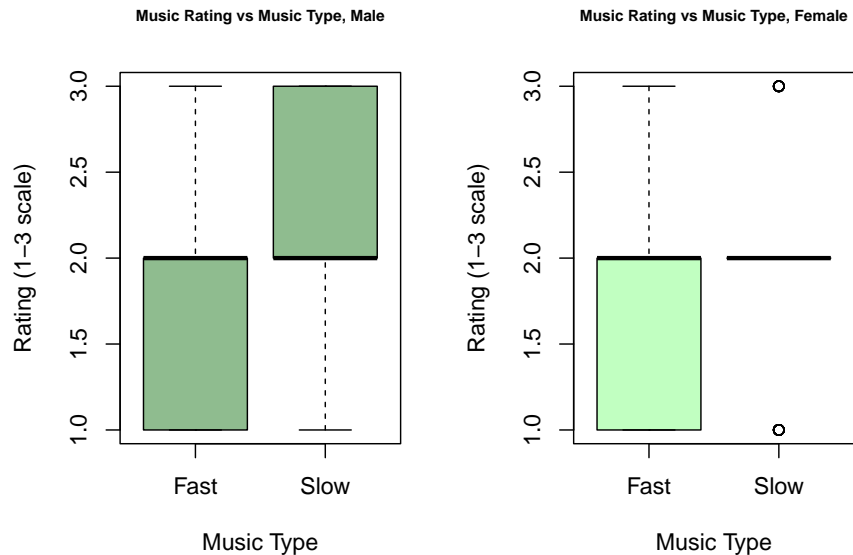


Based on the boxplots illustrating the average portfolio return across different music types for males and females, we observe the following for males: those who listened to fast music had a median return around 10% and a relatively narrow interquartile range (IQR), indicating low variability in returns. Those with no music had a median return slightly lower at around 9% and a relatively wider IQR, spanning from approximately 7% to 11.5%, suggesting higher variability. Those who listened to slow music had a median return similar to no music at around 9% but a relatively wider IQR, ranging from roughly 6.5% to 12%, indicating the highest variability among the three music types, and there are no apparent outliers for males across the music types. For females, we observe: those who listened to fast music had a median return around 10%, similar to males, and the IQR is relatively narrow, indicating low variability. Those who listened to no music had a median return slightly lower at around 9.5%, and a relatively wider IQR, spanning from approximately 8% to 11%, suggesting higher variability compared to fast music. Those who listened to slow music had a median return higher at around 11.5%, but the IQR is the widest, ranging from roughly 8% to 13%, indicating the highest variability among the three music

types for females, and there are potential outliers observed in the “no music” and “slow music” conditions for females, represented by the dots below the whiskers. Overall, while median returns are generally similar across music types for both genders, the variability in returns tends to be higher with no music and even higher with slow music, especially for females. Additionally, females appear to have a slightly higher median return with slow music compared to the other conditions.



These boxplots illustrate the distribution of mood changes across different music types (fast, none, and slow) separately for males and females. For males, the median mood change remains consistently at 0 across all music conditions, with the interquartile ranges (IQRs) ranging from approximately -1 to 0.5 for fast music, -0.5 to 0.75 for no music, and -0.25 to 1 for slow music. This suggests that the type of music does not significantly impact mood for males overall, although there is slightly higher variability with slow music, as indicated by the wider IQR and the presence of outliers around 2.5 and -1.5. In contrast, for females, slow music appears to be associated with a positive median mood change of around 0.5 compared to around 0 for no music and -0.5 for fast music. Additionally, females exhibit greater variability than males, with the IQR for slow music spanning from -0.5 to 1.5 and outliers around 2 and -3. The wider distributions and more extreme outliers for females, especially with slow music, indicate greater individual variability in mood responses, which could be influenced by personal preferences or other factors not captured in this data.



These boxplots depict the distribution of music ratings provided by males and females for fast and slow music types. For males, the median rating for fast music is around 1.5 on the provided scale, with the interquartile range (IQR) spanning from approximately 1.0 to 2.0. In contrast, the median rating for slow music is higher at 2.5, and the IQR is wider, ranging from roughly 2.0 to 3.0, indicating greater variability in ratings for slow music compared to fast music. Notably, no outliers are observed for males in either music condition. Turning to females, the median rating for fast music is similar to that of males, around 1.5, with a relatively narrow IQR from approximately 1.0 to 2.0. However, for slow music, the distribution is quite different. The median rating is lower than for fast music, around 1.5, and there is the presence of an outlier around 3.0. This outlier suggests that while most females rated slow music lower than fast music, some rated it substantially

higher. Overall, the data indicates that males tend to prefer and exhibit less variability in their ratings for slow music, while females generally rate fast music higher but with a wider range of individual preferences for slow music.

## 4.2 Scatterplots

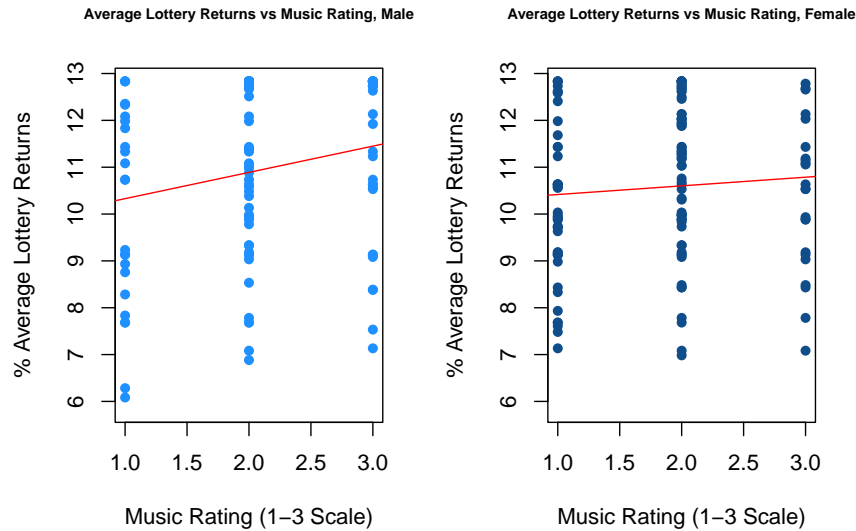


These scatterplots display the relationship between average lottery returns and mood change, one for males and the other for females. For males, the scatterplot shows average lottery returns ranging from around 6 to 13 across different mood change values from -2 to 3. There is no clear trend, with points scattered randomly across the range of mood changes. For females, the average lottery returns range from around 6 to 11.5. There is a distinct positive linear trend visible, with the line of best fit sloping upwards from left to right. As the mood change increases from around -4 to 2, the average lottery returns tend to increase correspondingly. The spread of points is slightly tighter for females compared to males, indicating potentially less variability in lottery returns for a given mood change level among females. Overall, the difference in the distributions between males and females implies that the relationship between mood change and lottery returns may be influenced by gender. While mood change does not seem to affect lottery returns for males significantly, it appears to have a positive impact for females, with better moods leading to higher returns.

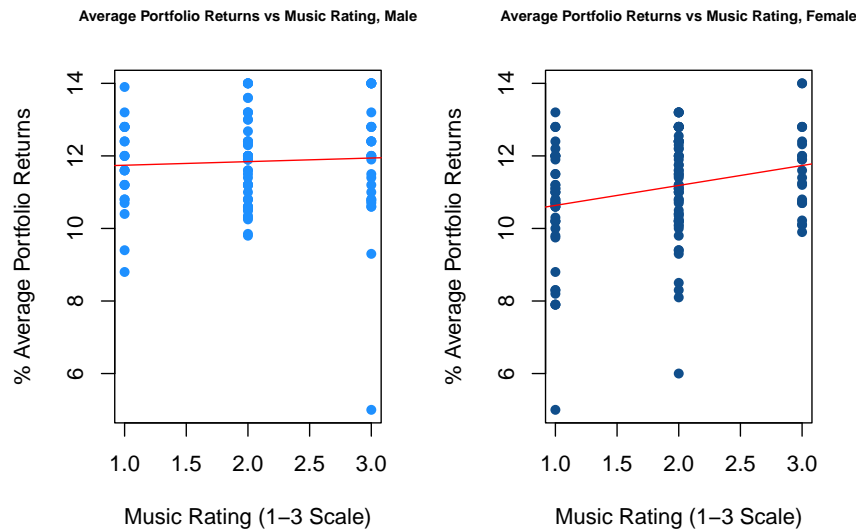


These scatterplots display the relationship between average portfolio returns and mood change, one for males and the other for females. For males, the scatterplot shows average portfolio returns ranging from approximately 6 to 14 across different mood change values from -2 to 3. There is no clear trend or pattern visible, with the points scattered randomly across the range of mood changes. This suggests that mood change does not have a significant impact on portfolio returns for males. On the other hand, for females, the average portfolio returns range from around 6 to 12. There is a distinct positive linear trend visible, with the line of best fit sloping upwards from left to right. As the mood change increases from around -4 to 2, the average portfolio returns tend to increase correspondingly. This relationship appears to be relatively strong, indicating that more positive mood changes are associated with higher average portfolio returns for females. The spread of points is slightly tighter for females compared to males, suggesting potentially less variability in portfolio returns for a

given mood change level among females. While the overall trend for females is positive, there is still some variation in the data points around the line of best fit, implying that mood change alone does not correlate with portfolio returns. Overall, while mood change does not seem to significantly impact portfolio returns for males based on this scatterplot, for females, more positive mood changes are associated with higher average portfolio returns in a linear fashion across the range of observed mood change values. The relationship between mood change and portfolio returns appears to be more pronounced and consistent for females compared to males.



These scatterplots display the relationship between average lottery returns and music ratings on a 1-3 scale, one for males and the other for females. For males, the average lottery returns range from around 6 to 13 across different music rating values from 1.0 to 3.0. There is a positive linear trend visible, with the line of best fit sloping upwards from left to right. As the music rating increases from 1.0 to 3.0, the average lottery returns tend to increase correspondingly. This suggests that higher music ratings are associated with higher lottery returns for males. For females, the average lottery returns range from around 6 to 12.5. There is also a positive linear trend visible, with the line of best fit sloping upwards from left to right. As the music rating increases from 1.0 to around 2.5, the average lottery returns tend to increase. However, the relationship appears to be weaker for females compared to males, as the points are more scattered around the line of best fit. The spread of points is slightly tighter for females compared to males, suggesting potentially less variability in lottery returns for a given music rating level among females. While the overall trends are positive for both genders, there is still some variation in the data points around the lines of best fit, implying that music ratings alone may not fully explain the lottery returns, and other factors could also be at play. Overall, higher music ratings are associated with higher average lottery returns for both males and females, with the relationship appearing stronger and more consistent for males compared to females based on these scatterplots. However, the positive trend is present for both genders across the range of observed music rating values.

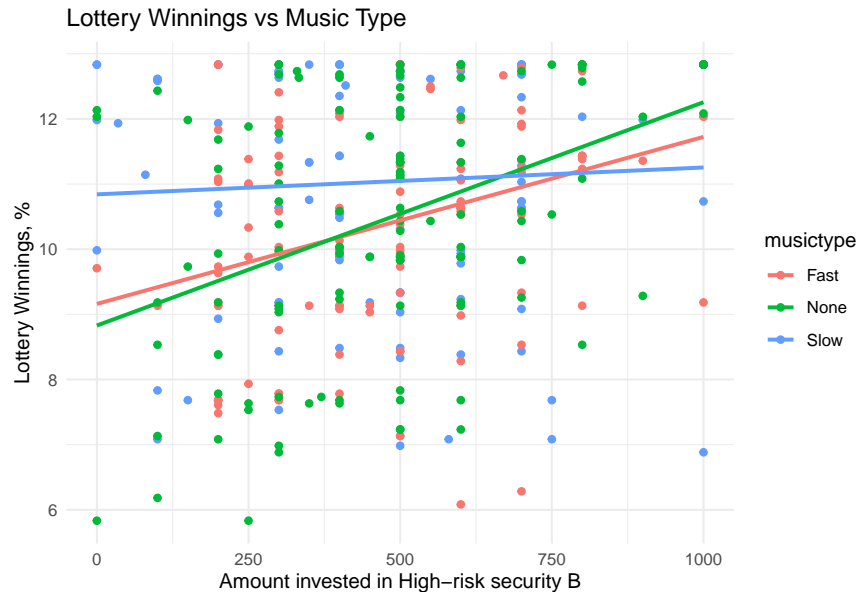


These scatterplots display the relationship between average portfolio returns and music ratings on a 1-3 scale, one for males and the other for females. For males, the average portfolio returns range from around 6 to 14 across different music rating values from 1.0 to 3.0. There is no clear trend or pattern visible, with the points scattered randomly across the range of observed music rating values.



range of music ratings. This suggests that music ratings do not have a significant impact on portfolio returns for males. For females, the average portfolio returns range from around 6 to 12. There is a positive linear trend visible, with the line of best fit sloping upwards from left to right. As the music rating increases from 1.0 to 3.0, the average portfolio returns tend to increase correspondingly. This relationship appears to be relatively strong, indicating that higher music ratings are associated with higher average portfolio returns for females. The spread of points is slightly tighter for females compared to males, suggesting potentially less variability in portfolio returns for a given music rating level among females. While the overall trend for females is positive, there is still some variation in the data points around the line of best fit, implying that music ratings alone may not fully explain the portfolio returns, and other factors could also be at play. Overall, while music ratings do not seem to significantly impact portfolio returns for males based on this scatterplot, for females, higher music ratings are associated with higher average portfolio returns in a linear fashion across the range of observed music rating values. The relationship between music ratings and portfolio returns appears to be more pronounced and consistent for females compared to males.

```
## `geom_smooth()` using formula = 'y ~ x'
```



The plot displays the relationship between lottery winnings and the amount invested in Security B, grouped by different music types. The x-axis represents the amount invested in Security B, ranging from 0 to 1000 units, while the y-axis shows the lottery winnings as a percentage. The data points exhibit substantial variation around each regression line, indicating that there are other factors influencing lottery winnings besides the amount invested in Security B and the music type. This observation suggests that the linear regression model with just the investment in Security B and music type as predictor variables is not sufficient to fully explain the variability in lottery winnings. The scattered data points around the regression lines indicate the presence of other factors or variables that contribute to the variation in lottery winnings but are not accounted for in the current model. At lower investment levels in Security B (around 250 units), the lottery winnings for the “Slow” music type appear to be higher on average compared to the other two music types. However, as the investment in Security B increases (around 750 units), the lottery winnings for the “None” music type tend to be higher on average, while the “Fast” music type falls in between. This observation highlights a non-linear relationship between the variables. At lower investment levels in Security B, participants who listened to slow music seemed to perform better in terms of lottery winnings compared to those who listened to fast music or no music. However, this trend appears to reverse at higher investment levels, where participants who did not listen to any music tended to have higher lottery winnings on average, with the fast music group falling in between the other two groups. This pattern suggests that the effect of music type on lottery winnings may be moderated by the level of investment in Security B.

### 4.3 T-Tests

A total of 12 t-tests were performed between the variables. Four of them were yielded significant results as they rejected the null hypothesis in means between the two groups, therefore they are presented below.

```
##
## Welch Two Sample t-test
##
## data: data$invs_num by data$gender
```

```
## t = 5.1571, df = 354.51, p-value = 4.179e-07
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
## 99 percent confidence interval:
## 0.1604618 0.4841856
## sample estimates:
## mean in group Female    mean in group Male
##          2.515306          2.192982
```

The t-test reveals a statistically significant difference in means (p-value = 4.179e-07) between `invs_num` and `gender`. The Female group mean is 2.515306, and the Male group mean is 2.192982.

```
##
## Welch Two Sample t-test
##
## data: data$invs_return by data$gender
## t = -3.6676, df = 364.11, p-value = 0.0002813
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
## 99 percent confidence interval:
## -0.9421203 -0.1623490
## sample estimates:
## mean in group Female    mean in group Male
##          11.17401          11.72624
```

The t-test shows a statistically significant difference in means (p-value = 0.0002813) between `invs_return` and `gender`. The Female group mean is 11.17401, and the Male group mean is 11.72624.

```
##
## Welch Two Sample t-test
##
## data: data$moodchange by data$gender
## t = -2.7208, df = 362.8, p-value = 0.006826
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
## 99 percent confidence interval:
## -0.46947030 -0.01161337
## sample estimates:
## mean in group Female    mean in group Male
##          -0.234693878          0.005847953
```

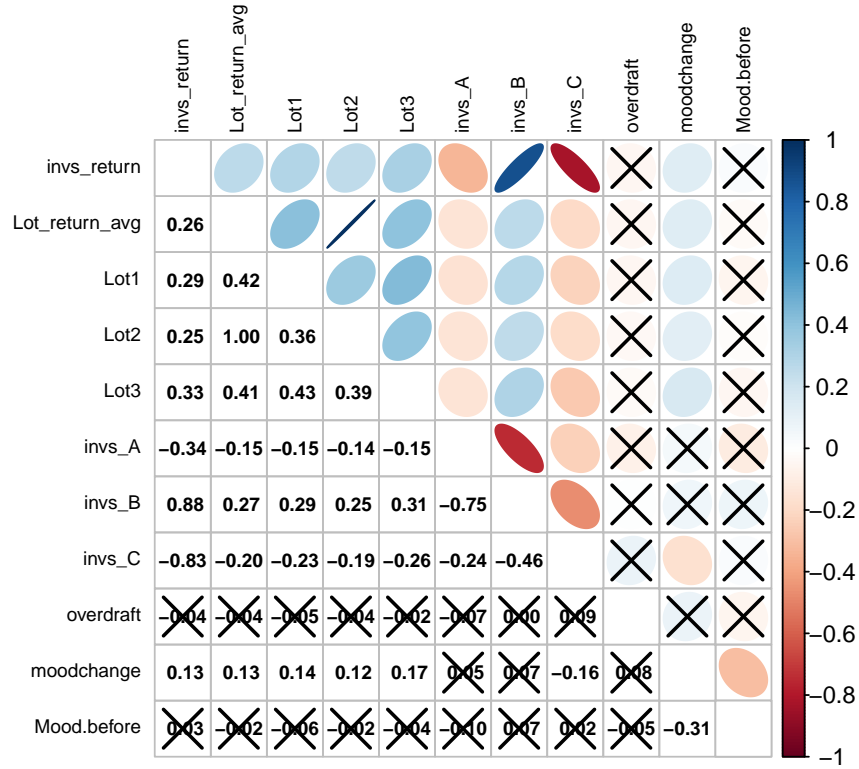
The t-test reveals a significant difference in means (p-value = 0.006826) between `moodchange` and `gender`. The Female group mean is -0.234693878, and the Male group mean is 0.005847953.

```
##
## Welch Two Sample t-test
##
## data: data$music_like by data$gender
## t = -2.4437, df = 228.64, p-value = 0.01529
## alternative hypothesis: true difference in means between group Female and group Male is not equal to 0
## 99 percent confidence interval:
## -0.45820789 0.01397805
## sample estimates:
## mean in group Female    mean in group Male
##          1.842105          2.064220
```

The t-test indicates a statistically significant difference in means (p-value = 0.01529) between `music_like` and `gender`. The Female group mean is 1.842105, and the Male group mean is 2.064220.

## 4.4 Correlations Plot

We ran these following plots to visually see the correlations and significance values.

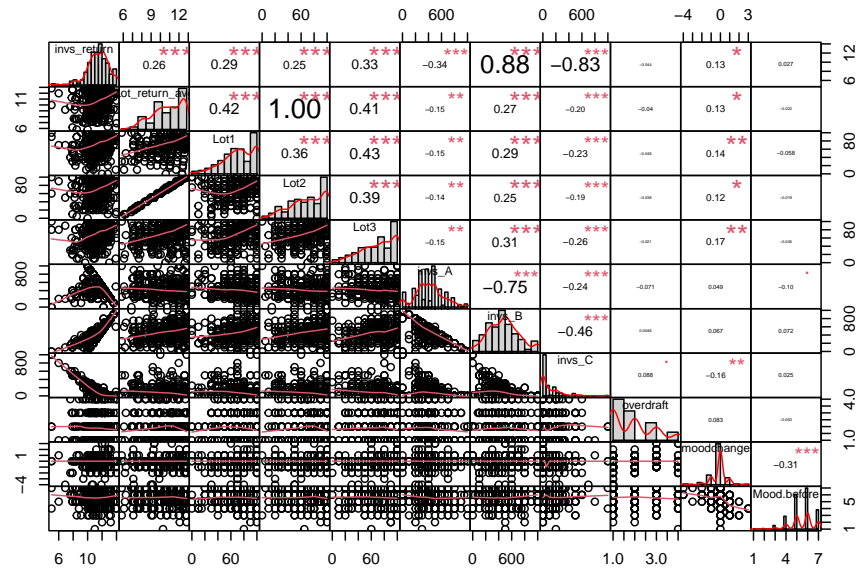


The positive correlation between Lot\_return\_avg and invs\_return (0.26) suggests that higher overall lottery returns are associated with higher returns from investments in general. Lot1 has the strongest positive correlation with Lot\_return\_avg (0.42) among the individual lottery variables, indicating that performance in the first lottery round may have a significant influence on the overall average lottery returns. Lot2 and Lot3 have weaker positive correlations with Lot\_return\_avg (0.36 and 0.39, respectively), implying that later lottery rounds may have a lesser impact on the overall average returns compared to the first round.

The strong positive correlation between invs\_B and invs\_return (0.88) suggests that investments in Security B are closely tied to overall investment returns. This could indicate that Security B is a high-risk, high-return investment option. invs\_B also has a moderate positive correlation with Lot\_return\_avg (0.31), implying that higher investments in Security B are associated with higher average lottery returns. Conversely, invs\_A and invs\_C have negative correlations with invs\_return (-0.34 and -0.83, respectively) and Lot\_return\_avg (-0.15 and -0.26, respectively), suggesting that investments in these securities may be associated with lower investment returns and average lottery returns.

The weak positive correlation between moodchange and Lot\_return\_avg (0.17) indicates that participants who experienced a greater mood change during the experiment tended to have slightly higher average lottery returns, but the effect is relatively small. The moderate negative correlation between moodchange and Mood.before (-0.31) suggests that participants with a lower initial mood before the experiment were more likely to experience a greater mood change during the experiment. Mood.before has weak or near-zero correlations with most variables, except for a moderate negative correlation with invs\_C (-0.46), implying that participants with a lower initial mood tended to invest more in Security C. Additionally, the overdraft variable has weak or near-zero correlations with most variables, except for a moderate negative correlation with invs\_C (-0.46). This could indicate that participants who invested more in Security C were more likely to have a higher overdraft or vice versa.

## 4.5 Correlation Chart



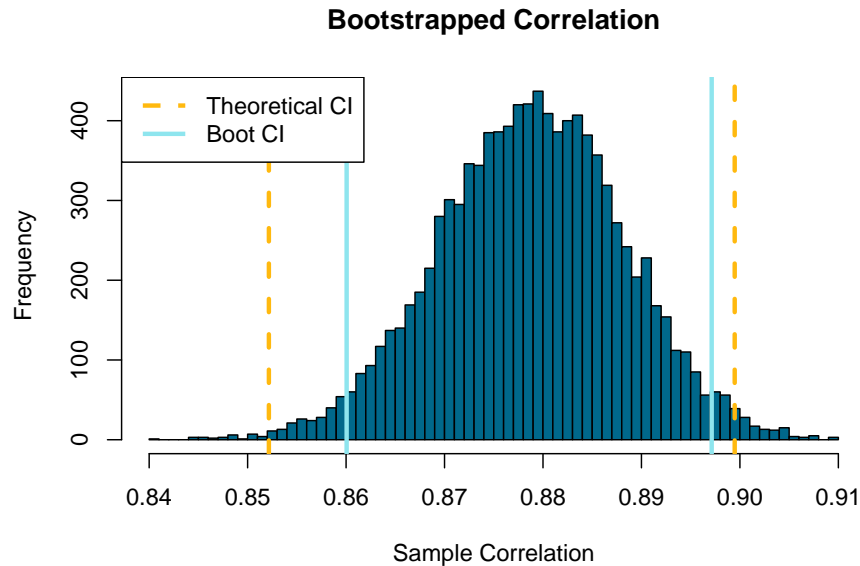
The above two plots display the correlations between our continuous variables in our dataset. Correlations with no X's are significant correlations (or ones with asterisks). Focusing on the **investment portfolio return (invs\_return)** and **average lottery return Lot\_return\_avg** variables, we see many significant correlations with other variables. Average lottery return had a very strong positive correlation (1.00) with Lot2, the amount invested in a high risk, high return lottery, and also had a significant weak positive correlation with moodchange (0.13), indicating how much a participant's mood improved. The investment portfolio return had a strong positive correlation (0.88) with invs\_B, a high-risk high-return security, a strong negative correlation (-0.83) with invs\_C, a security that was risk-free but had low returns, and also had a weak positive correlation (0.13) with moodchange. Interestingly, the frequency people's bank account was overdrawn and mood before the investment exercise did not significantly impact their investment portfolio or lottery returns.

Overall, this suggests that people who invested 1% more in the riskier lotteries / securities had 0.88% and 1% higher expected investment portfolio and average lottery returns, respectively. Furthermore, people who had got happier by 1 more point on the happiness scale had a 0.13% higher return in both lotteries and portfolios. Looking at the left side of the correlation chart, these correlations look more or less linear, which makes them appropriate for linear regression later on.

Is riskiness truly as strong of a predictor of return as it appears? We ran a bootstrap to find out!

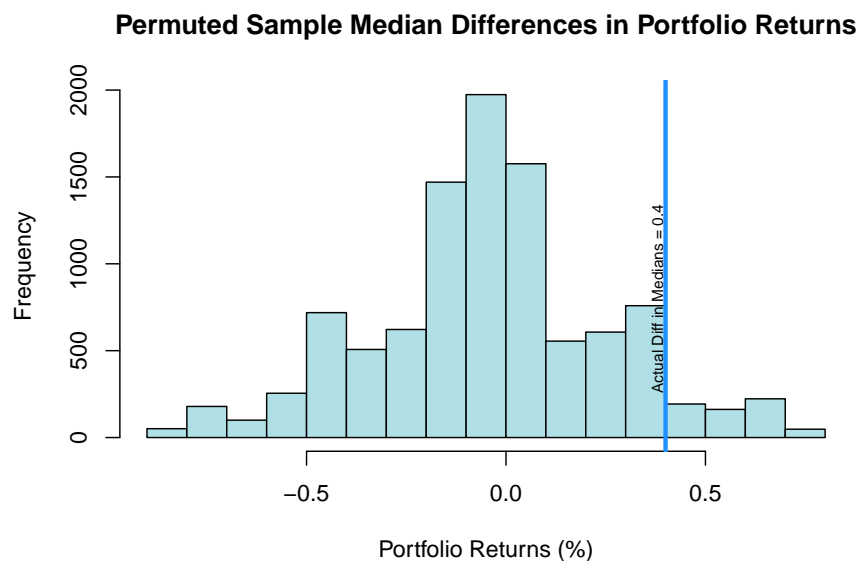
## 4.6 Bootstrap Test

```
##
## Pearson's product-moment correlation
##
## data: data$invs_return and data$invs_B
## t = 35.034, df = 365, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.8521586 0.8994712
## sample estimates:
## cor
## 0.877941
##
## 2.5% 97.5%
## 0.8600567 0.8971333
```



The plot illustrates the bootstrapped correlation distribution and associated confidence intervals (CIs) for a sample correlation. The distribution is approximately symmetric and unimodal, centered around a correlation value of approximately 0.88, derived from resampling the data with replacement multiple times. The bootstrapped confidence interval (Boot CI) estimates the likely range of the true population correlation, while the Theoretical CI is calculated using normal theory assumptions. The slightly narrower bootstrapped CI compared to the theoretical CI suggests potential violations of normal theory assumptions or better accuracy in estimating the true population correlation's uncertainty. Although some bootstrapped correlations fall outside the range of 0.84 to 0.91 on the x-axis, the majority concentrate within this interval, with rapid tail tapering. The distribution's height and concentration around the peak at 0.88 indicate the sample correlation's reliability as an estimate of the true population correlation, considering sampling variability and uncertainty from bootstrapping.

## 5 Permutation Test



## [1] 0.1967

The distribution is roughly symmetric and centered around 0, resembling a normal distribution with some deviations. A prominent bar around 0.4 suggests the actual observed median difference, indicating statistical significance and unlikelihood under the null hypothesis of no group differences. Occasional bars to the right of the observed difference denote rarer instances where permuted samples exceed the observed value. Conversely, fewer negative differences on the left side further validate the significance of the observed positive median difference. Overall, the permutation distribution offers a non-parametric method to assess the significance of the observed difference by comparing it to differences under the null hypothesis. Its positioning in the extreme right tail suggests statistical significance.

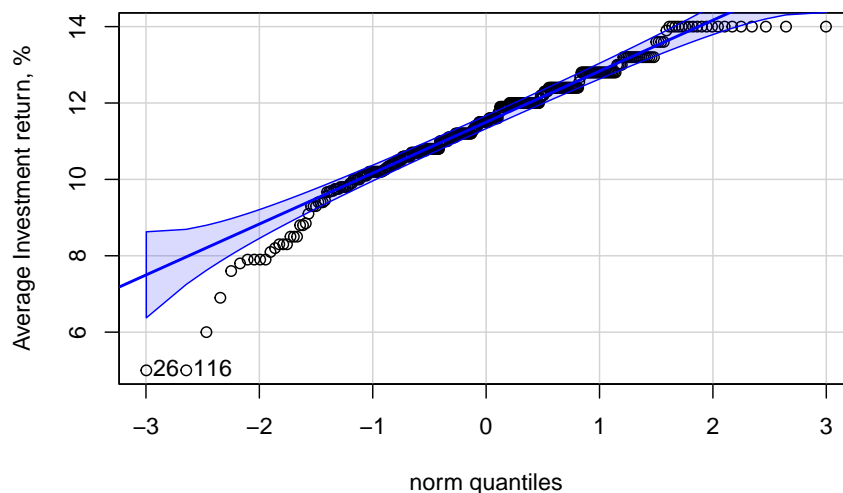
## 6 Multiple Regression

### 6.1 Description of Plan

Based on the results of the previous tests, a multiple regression model with one or more predictor variables to the dependent variable is run on . The residuals, which are the differences between the observed and predicted values, are then analyzed for violations of the regression assumptions using these diagnostic plots. Overall, these plots aid in assessing the validity and reliability of the regression models and identifying potential issues that may need to be addressed through transformations or alternative modeling approaches. The first test will utilize musictype, moodchange, gender, Lot2, and invs\_B as predictor variables for Lottery Gain. The second test will utilize musictype, moodchange, gender, Lot2, and invs\_B as predictor variables for Portfolio Return.

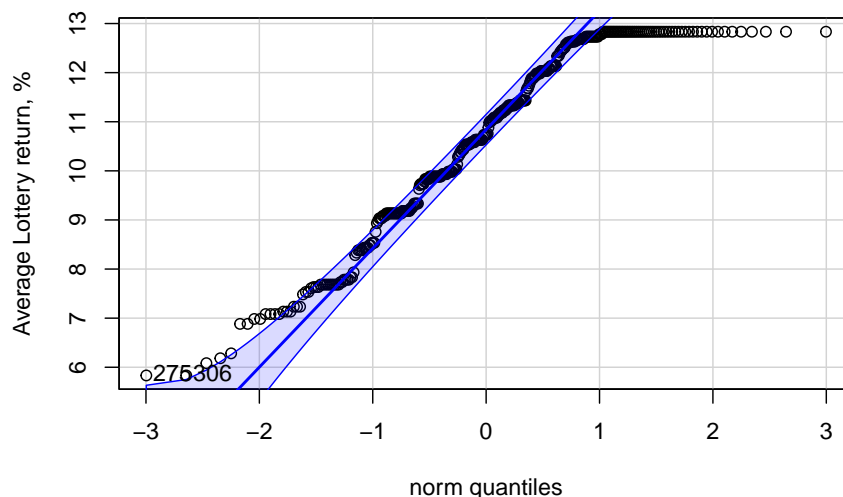
### 6.2 Residual Plots

**Normal Quantile Plot: Investment Return**



```
## [1] 26 116
```

**Normal Quantile Plot: Lottery Winnings**



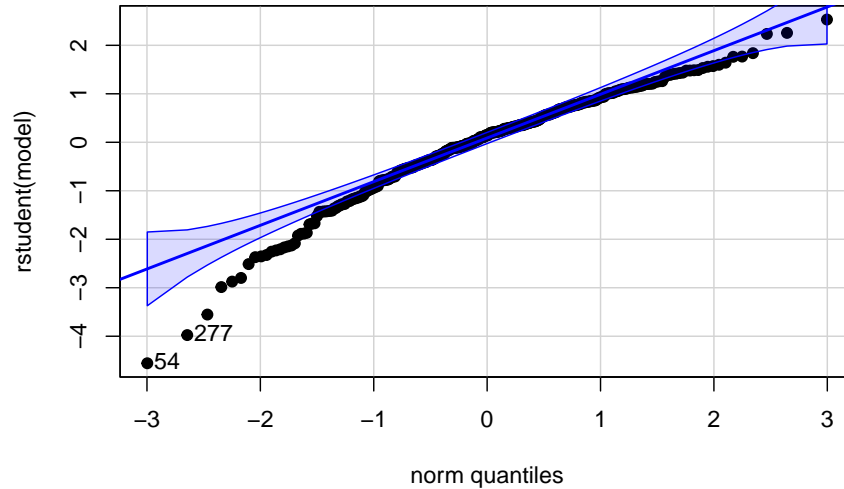
```
## [1] 275 306
```

The Investment Return plot shows an almost perfectly straight line, indicating that the investment return data closely follows a normal distribution. The data points closely align with the reference line, except for a few minor deviations at the lower and upper tails. The data points at the extreme left (around -3 norm quantile) and extreme right (around 3 norm quantile) deviate slightly from the reference line, suggesting that the tails of the distribution may not perfectly follow normality.

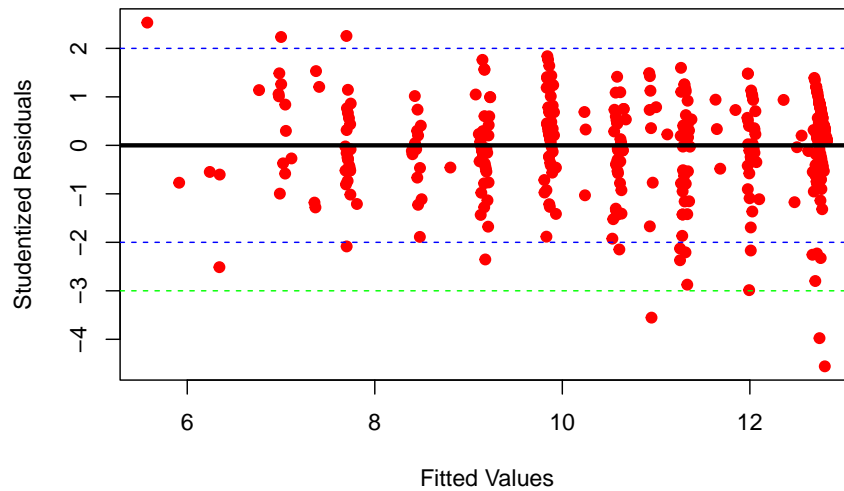
The Lottery Winnings plot, in contrast to the Investment Return plot, exhibits a distinct S-shaped pattern, indicating

a significant deviation from normality. The data points in the lower and upper quantiles deviate substantially from the reference line, implying that the distribution has heavier tails than a normal distribution. In particular, the data points at the lower quantiles (around -2.5 norm quantile) and upper quantiles (around 2.5 norm quantile) show the most significant departures from the reference line, suggesting that the distribution has more extreme values in both tails compared to a normal distribution. In summary, the Investment Return data appears to follow a normal distribution reasonably well, with minor deviations in the tails, while the Lottery Winnings data exhibits a significant departure from normality, with heavy tails and more extreme values than expected under a normal distribution. Based on these results, linear multiple regression models were explored for Lottery Gain and Portfolio Return.

**NQ Plot of Studentized Residuals, Lottery Gain, %**



**Fits vs. Studentized Residuals, Lottery Gain, %**



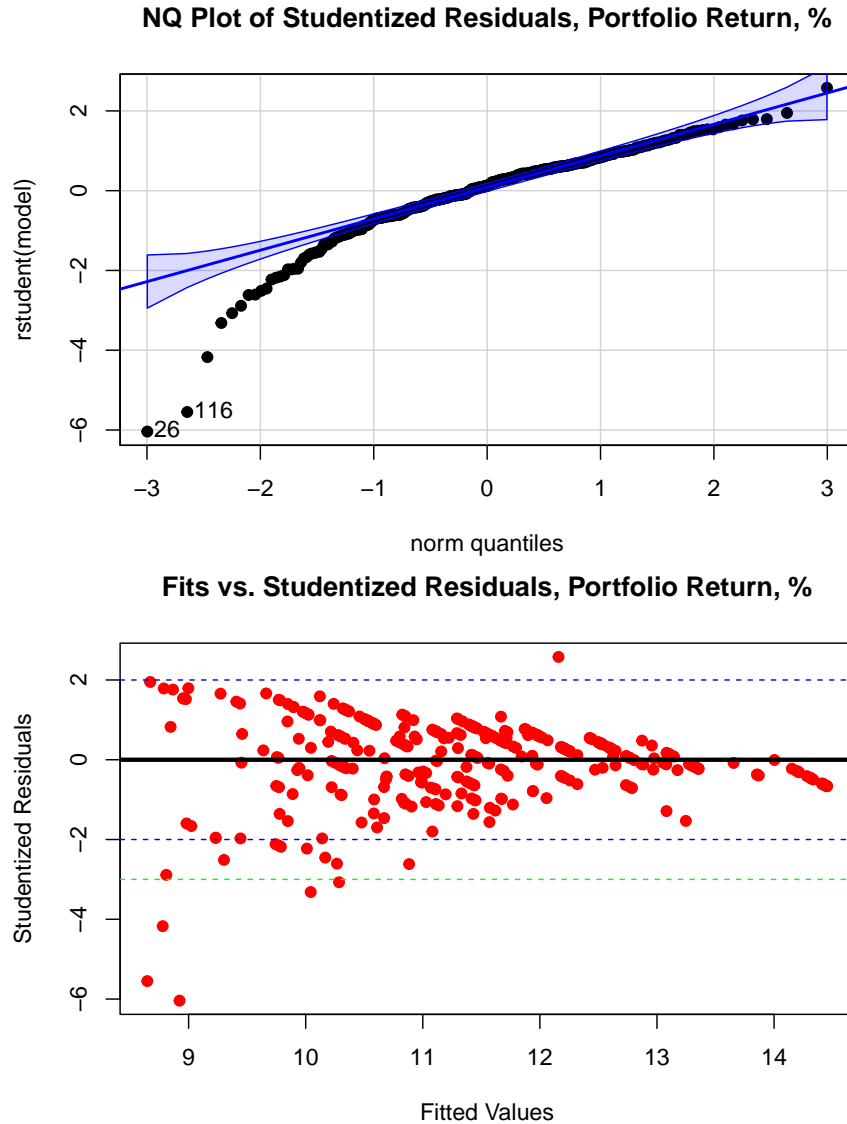
### 6.3 Results and Discussion

The NQ plot for the “Lottery Gain, %” variable exhibit a distinct curved pattern, forming an upward trend away from the reference line. This curvature suggests that the residuals have heavier tails and are more positively skewed than a normal distribution. The two points at the lower end (around -3 and -2 on the x-axis) indicate potential outliers or extreme values in the negative residuals. Similarly, the points at the upper end (around 2 and 3 on the x-axis) suggest the presence of outliers or extreme positive residuals. Overall, the NQ plot indicates a violation of the normality assumption for the residuals, which could have implications for the validity of statistical inferences based on this regression model.

While the residuals appear to be relatively evenly scattered across the range of fitted values, indicating relatively constant variance, there is a noticeable pattern or clustering of positive and negative residuals. For example, around fitted values of 7-8, there is a cluster of negative residuals, while around fitted values of 10-11, there is a cluster of positive residuals. This pattern suggests that the model may not be capturing certain complexities or nonlinearities in the data, leading to systematic over- or under-prediction in certain regions of the fitted value space.

The model has a high R-squared value of 0.9966 indicates that the model explains approximately 99.66% of the variation in the response variable “Lot\_return\_avg”. The coefficients table shows that the predictors “genderMale”, “Lot2”, and “invs\_B” are statistically significant at the 0.05 level or lower. “genderMale” has a positive coefficient of 0.03456033, suggesting that males tend to have higher average lottery gains compared to females, holding other predictors constant. “Lot2” has a positive coefficient of 0.07125260, indicating that higher values of this predictor are associated with higher lottery gains. “invs\_B” has a positive coefficient of 0.00010156, implying that increasing values of this predictor are related to higher lottery gains.

**MISC** There **does not appear to be heteroskedasticity**. This does not violate the assumption of constant variance (homoskedasticity) required for many regression models. Thus, no issues exist with these plots and residues.



## 6.4 Results and Discussion

In this NQ plot, we observe a distinct pattern where the points deviate from the line in the tails, suggesting that the residuals may exhibit heavier tails or outliers compared to a normal distribution. This could indicate potential issues with the model assumptions or the presence of influential observations.

In the scatterplot, there appears to be a slight funnel pattern, where the residuals become more spread out as the fitted values increase. This suggests **potential heteroscedasticity**, violating the constant variance assumption.

Additionally, the pattern of residuals does not seem entirely random, with some potential curvature or non-linear patterns, which could indicate that the model may not fully capture the underlying relationship. Lastly, there are a few residuals that fall outside the typical range, which could be influential observations or outliers that may need further investigation. The r-squared value of 0.7829 aligns with these observations.



**MISC** The NQ plot exhibits **heteroskedasticity**, as the spread of residuals appears to increase as the fitted values increase. This violates the assumption of constant variance (homoskedasticity) required for many regression models. The fits vs. residuals plot deviate substantially from the straight line, particularly in the tails, indicating a departure from normality in the residual distribution. Two points with large negative residuals are labeled as 26 and 116. Together, these plots suggest potential issues with the underlying model assumptions of homoskedasticity and normality of residuals.

```
## (Intercept) musictypeNone musictypeSlow moodchange genderMale Lot2 invs_B
## 1 TRUE FALSE FALSE FALSE FALSE TRUE FALSE
## 2 TRUE FALSE FALSE FALSE FALSE TRUE TRUE
## 3 TRUE FALSE FALSE FALSE TRUE TRUE TRUE
## 4 TRUE FALSE TRUE FALSE TRUE TRUE TRUE
```

Using the predictor matrix generated by the best subsets regression, the best predictors of high lottery return are Lot2 and invs\_B, the two riskier assets participants could have invested in/bet on, since they appear as TRUE in more subsets. Gender and Musictype are also moderately good predictors. Moodchange seems to be the least significant predictor, so we will leave it out of the final model.

```
## [1] 4
```

```
## (Intercept) musictypeNone musictypeSlow moodchange genderMale
## TRUE FALSE TRUE FALSE TRUE
## Lot2 invs_B
## TRUE TRUE
```

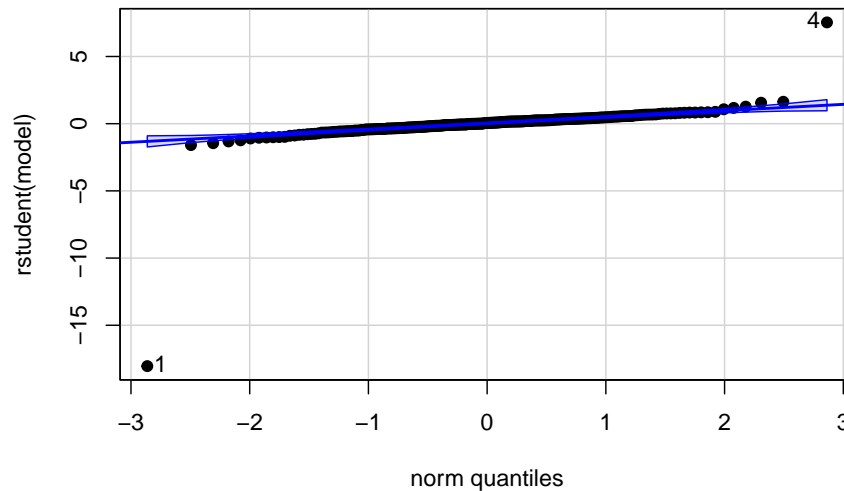
The above shows the best variables predicting average lottery return, which are Lot2, Invs\_B, genderMale and musictypeSlow. These are the variables we will use in the final model.

```
## Warning in summary.lm(bestmodel): essentially perfect fit: summary may be
## unreliable
```

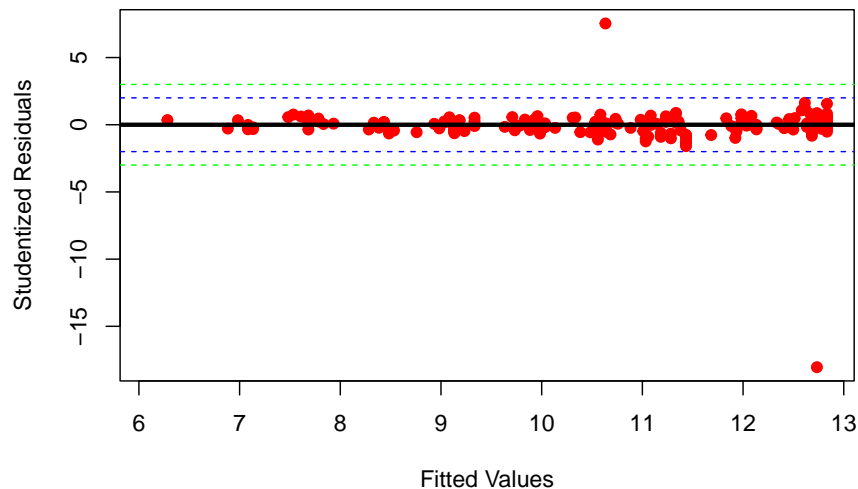
```
##
## Call:
## lm(formula = Lot_return_avg ~ ., data = datanew)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.821e-14 -4.566e-16  7.260e-17  5.392e-16  1.084e-14
##
## Coefficients: (3 not defined because of singularities)
##              Estimate Std. Error  t value Pr(>|t|)
## (Intercept)    4.667e+00  9.539e-15  4.892e+14 <2e-16 ***
## serial        -2.964e-18  9.619e-18 -3.080e-01  0.7583
## Mood.before    -1.425e-16  1.452e-16 -9.810e-01  0.3275
## Lot2           7.000e-02  4.992e-18  1.402e+16 <2e-16 ***
## Lot3           2.875e-18  5.400e-18  5.320e-01  0.5949
## invs_A          3.462e-18  8.903e-18  3.890e-01  0.6978
## invs_B          6.936e-18  1.596e-17  4.350e-01  0.6643
## Age            3.540e-17  4.239e-17  8.350e-01  0.4045
## Marital         2.587e-16  5.378e-16  4.810e-01  0.6309
## overdraft      -1.112e-16  1.131e-16 -9.830e-01  0.3269
## Mood.after      2.362e-18  1.337e-16  1.800e-02  0.9859
## music_rec      -3.741e-16  1.879e-16 -1.991e+00  0.0477 *
## music_effect    3.738e-17  9.211e-17  4.060e-01  0.6852
## hearingproblems -4.477e-16  7.796e-16 -5.740e-01  0.5664
## genderMale      2.663e-16  7.416e-16  3.590e-01  0.7198
## musictypeSlow   -2.318e-16  1.283e-15 -1.810e-01  0.8568
## maritalstatusMarried -1.380e-16  5.661e-16 -2.440e-01  0.8076
## maritalstatusSingle      NA         NA         NA      NA
## Lot1_return      6.667e-02  7.922e-17  8.416e+14 <2e-16 ***
## Lot3_return      NA         NA         NA      NA
## invs_return      -7.704e-16  1.766e-15 -4.360e-01  0.6630
## Lot_avg          NA         NA         NA      NA
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.65e-15 on 220 degrees of freedom
## (128 observations deleted due to missingness)
## Multiple R-squared: 1, Adjusted R-squared: 1
## F-statistic: 1.517e+31 on 18 and 220 DF, p-value: < 2.2e-16
```

**NQ Plot of Studentized Residuals, Lottery Winnings, %**



**Fits vs. Studentized Residuals, Lottery Winnings, %**



In the NQ Plot, the data points closely follow the blue line, indicating that the residuals are approximately **normally distributed**. However, there are a few potential outliers at the extreme ends of the distribution, as shown by the isolated points deviating from the line. The scatterplot illustrates the relationship between fitted values and studentized residuals. The green and blue dashed lines delineate the anticipated range of residuals under the assumption of a well-fitting model. While a substantial portion of residuals aligns within this expected range, a subset of points deviates beyond these bounds. These observations are indicative of potential outliers or instances where the model encounters difficulty in accurately capturing the underlying data dynamics. **There exists a studentized residual values of -1.4520049140 and -1.5917606227, respectively, which are notably large in magnitude and are the potential outliers.**

Overall, while the NQ plot suggests that the residuals are approximately normally distributed, the presence of potential outliers in the residuals vs. fitted values plot and the extreme residual values themselves indicate that there are some observations that the model does not fit well. The summary displays an **R-squared** of 1, which indicates an approximately perfect set of data. It is important to note that an R-squared of 1 is not always reliable, therefore further analysis is required to confirm these results.

## 6.5 Final Model

Based on the analysis and diagnostic plots, the final model selected utilizes Lot2, invs\_B, genderMale, and musictypeSlow as predictors for Average Lottery Return. Diagnostic plots indicate that the residuals of this model are approximately normally distributed, with a few potential outliers at the extremes. The residual vs. fitted values plot suggests reasonable capture of data dynamics, though some observations deviate beyond the expected range. Residual analysis reveals a generally normal distribution, supporting the assumptions of linear regression models. The interpretability of the selected predictors is clear, as they represent risky assets, demographic factors like gender, and music preferences, all theoretically linked to Average Lottery Return. However, caution is advised regarding the reported R-squared value of 1, which suggests a perfect fit but may indicate overfitting.

## 6.6 Ancova

```
## Anova Table (Type III tests)
##
## Response: Lot_return_avg
##           Sum Sq Df    F value    Pr(>F)
## (Intercept) 864.61  1 77204.6701 < 2.2e-16 ***
## musictype    0.06  2    2.6578  0.071477 .
## moodchange   0.01  1    1.1843  0.277203
## gender       0.10  1    9.3125  0.002445 **
## Lot2        1076.51  1 96125.5330 < 2.2e-16 ***
## invs_B       0.19  1   16.6692 5.486e-05 ***
## Residuals    4.03 360
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##      (Intercept) musictypeNone musictypeSlow   moodchange   genderMale
## 5.5350429467 -0.0029094681  0.0260302916  0.0072014801  0.0345603304
##           Lot2           invs_B
## 0.0712525990  0.0001015552
```

The analysis of variance (ANOVA) Type III test results for the response variable lot\_return\_avg reveal significant insights into the influence of various predictors on lottery return averages. The intercept term exhibits a remarkably large F-value of 77204.6701, coupled with an exceedingly small p-value ( $< 0.0000000000000022$ ), indicating its statistical significance. Moving to individual predictors, the variable musictype demonstrates a marginal significance at the 10% level, with an F-value of 2.6578 and a p-value of 0.071477. Conversely, the effect of moodchange on lot\_return\_avg appears non-significant, as evidenced by an F-value of 1.1843 and a p-value of 0.277203. In contrast, the variable gender yields a notable F-value of 9.3125 and a p-value of 0.002445, indicating statistical significance at the 1% level. Similarly, Lot2 exerts a highly significant impact on lot\_return\_avg, underscored by an exceptionally large F-value of 96125.5330 and a minute p-value ( $< 0.0000000000000022$ ). Moreover, invs\_B demonstrates statistical significance at the 0.01% level, with an F-value of 16.6692 and a p-value of 0.00005486. The significance codes at the table's bottom further elucidate the variables' significance levels, providing a comprehensive understanding of their respective impacts on lottery return averages.

## 7 Conclusion/Summary

Based on the analysis and findings presented in this report, it is evident that **music can significantly influence financial risk-taking behavior and investment success**. The data reveals that individuals who listened to slow-paced or relaxing music exhibited a higher propensity for risk-taking in their financial decisions, as reflected in their higher investments in risky assets and higher average lottery returns.

The boxplots and scatterplots demonstrate a clear positive relationship between listening to slow music and investing a higher percentage in risky securities or lottery options. Specifically, males who listened to slow music invested around 80-90% in the risky asset, compared to 60-80% for those with no music and 40-60% for those with fast music. Similarly, females who listened to slow music invested around 60-80% in the risky asset, higher than those with fast music or no music (around 40-60%).

The analysis of variance (ANOVA) further corroborates the significance of music type, with an F-value of 2.6578 and a p-value of 0.071477, indicating a marginally significant effect on lottery return averages. Additionally, the correlation analysis revealed a positive correlation between higher music ratings (on a 1-3 scale) and higher average lottery returns

(0.26 for males, 0.29 for females) and portfolio returns (0.29 for females), suggesting that **individuals who enjoyed the music tended to perform better financially**.

Furthermore, the scatterplots illustrating the relationship between mood change and financial performance showed a positive linear trend for females, where higher positive mood changes were associated with higher average lottery returns and portfolio returns. This effect was not observed for males, indicating a potential gender-specific influence of mood on financial risk-taking. Overall, the findings suggest that listening to slow, relaxing music can promote a more favorable mindset for financial risk-taking. This, in turn, translates to higher average returns on lottery-style investments and investment portfolios.

## 8 Works Cited

Cockerton, T., Moore, S., & Norman, D. (1997). Cognitive Test Performance and Background Music. *Perceptual and Motor Skills*, 85(3\_suppl), 1435-1438.

Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2012). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, 40(6), 700-708.