

# **An Analysis of Attributes that Contributed to The Cure Rock Band's Popularity**

PROJECT REPORT SUBMITTED  
IN FULFILMENT OF THE REQUIREMENTS FOR THE COURSE  
STAT 250 – APPLIED STATISTICS

DEPARTMENT OF STATISTICS OF  
MIDDLE EAST TECHNICAL UNIVERSITY

BY

Group 6:

Abtin Mogharabin - 2531739

Ali Muhammad Lalani - 2531135

Abdul Samad Najm - 2531242

July 2022

## **ABSTRACT**

Many rock bands, over the decades, have influenced Rock-Music. The Cure rock band has been classified among the most influential rock bands. Their creativity and virtuosity in music allowed them to rise to fame more quickly than other rock bands. In this research, we analyzed the data taken from Kaggle to investigate the central theme behind

- The attributes that allowed The Cure rock band to rise in fame.

The results outlined that most of the band's popular songs were euphoric and ecstatic rather than sad and downbeat. Their hit songs were more major scale than minor scale. Furthermore, their three influential albums were more energetic. Last but not least, true to what most fans perceive, the Cure rock band's famous songs had a loud threshold to them.

## **1. Introduction**

From riches to rags, it has been an ironic journey for the Cure Band. From being at the top of the music world and making it into the billboard charts week in week out, to the state where no one prioritizes to listen their songs. It is quite hard to digest how a band so successful and thriving came down to hit the rock bottom. But as far as their achievements are concerned, they were an exceptional band in their peak years, and this is what all of us collectively agree upon. We believe that their discography was something totally out of this world and well ahead of its time. There is absolutely no doubt why we all consider the Cure Band as one of our most favorite music groups and that is why we have based our research on it. Their songs are so relatable that they directly reach your heart. We always wanted to explore the reasons behind their success and their qualities because of which millions and millions of people listen to their songs on everyday basis. We wanted to know what makes them so special, what's their niche and why they were able to win so many hearts with their back-to-back hit albums in such a short period of time. To analyze this, we selected a dataset from Kaggle which epitomizes all of the important features and characteristics their songs had. We have variables like track popularity, loudness and energy etc. which gives some valuable insight on the songs that the Cure Band produced over the years. The variables that we have specifically used from our dataset are track popularity, danceability, liveness, valence, loudness, tempo, energy, instrumentalness, and key mode. These variables helped us evaluate and dive deep in the

discography. For example, we used track popularity to identify the most popular albums as it tells how popular an album was on a scale of 0 to 100. Similarly, we've also incorporated valence variable in our research which is a measure of positivity a track possesses between 0 and 1. Moreover, the energy variable that ranges from 0 to 1 was also used to interpret the intensity and activity of each track in their albums. We also used loudness and key mode to examine whether the music was loud or not on a decibel scale (-60 to 0 db) and to know the major and minor keys respectively, on which a song is composed. Other than that, danceability and tempo were used to study how suitable a track is to dance on and the overall speed of the beats and rhythm a song follows. Finally, the liveness variable was also taken to detect if a track was recorded live or not. In this analysis, we have used all these variables to study and assess the music that the Cure Band produced in such a way that helped us realize what key features helped them rise to fame in such a short time. This was our aim for the study, to try our best to discover the features in their songs that made people go crazy. For our research we have used different statistical methods to prove some amazing claims. From Analysis of Variances (One-way ANOVA) and the non-parametric Kruskal–Wallis's test to Multiple and Simple linear regression analysis, we have implemented the best statistical procedures to establish our conclusions. Furthermore, we also used one and two sample t-tests and one and two proportion z-tests as well. Throughout our research, we have made sure to satisfy all the assumptions first and then only apply the specific tests to prove our hypotheses. All these efforts we undoubtedly made to investigate more and more about the success of the band that we all love and admire. The Cure Band holds a special place in all of our hearts, and it was high time that they received some appreciation for the exceptional work they have done which is now beautifully expressed in our research project.

### **1.1.Data description**

In this project, we used the data collected by Xavier on Kaggle. This dataset contains 224 observations and 23 variables and could be found on <https://www.kaggle.com/datasets/xvivancos/the-cure-discography>. Each observation in our data is one of the tracks played by The Cure Band from 1982 to 2011. Our variables also include 9 continuous variables, 5 discrete variables, 9 categorical variables, and 2 time series variables. Here we can see a summary of the important variables of our dataset:

<b>Variables</b>	<b>Description</b>
album_name	The name of album for each track
album_release_date	Release date of each album
track_name	Name of the track
danceability	How suitable the track is to dance with – includes a value between 0.0 and 1.0
energy	A measure of fastness, loudness, and intensity of each track - includes a value between 0.0 and 1.0
key	The estimated key used for the track in general
loudness	Loudness of the track – mostly ranged between -60 and 0 db
mode	Shows modality of the track (major/minor)
speechiness	Shows the presence of spoken words in the track - a value between 0.0 and 1.0
acousticness	Shows whether the track is acoustic or not - a value between 0.0 and 1.0
instrumentalness	Predicts whether the track contains no vocals - a value between 0.0 and 1.0
liveness	Detects the existence of audience when recording the track - a value between 0.0 and 1.0
valence	Shows the positiveness of the track (happy/cheerfulness) - a value between 0.0 and 1.0
tempo	Shows the speed that the track is being played at
duration_ms	Shows the duration of the track in milliseconds
time_signature	Shows overall time signature of the track <i>Note: The time signature is a measure to specify how many beats are in each bar</i>
key_mode	Showing the combined result of both “key” and “mode” variables
track_popularity	The value showing the popularity of the track worldwide – an integer between 0 and 100

## 1.2. Research questions

We had 6 main questions analyzed in our study:

- 1) What factors are correlated with Track-Popularity?

- 2) When was the cure band more successful? When using major scale music or minor scale music?
- 3) Is there any statistically significant difference between the mean energy levels of the albums in the three most popular albums at 0.05 significance level?
- 4) Can we claim that the instrumentalness of the songs is also effective on how the track effects the audience for the tracks in question 3?
- 5) In modern rock music, the famous bands usually create song which have a high loudness. Was this factor also important in The Cure band's music?
- 6) It is usually considered by people that, in the rock genre more popular tracks are played at a high speed. Is this also a correct for The Cure?

### **1.3. Aim of the study**

In this study, we analyze the main reasons of The Cure band's fame. It is known that The Cure band started their special music style without having much expectations from it. But in short amount of time, they became one of the most known bands of their time. Our main aim here is to see what are the main factors that effected the audience the most. What was so special about The Cure band's that makes them still so famous even after 30 years passing from the time of their popularity?

## **2. Methodology/Analysis**

After deciding on the aim of our study and research questions, we started to choose suitable statistical analysis methods. For our analysis we used the suitable functions in R programming. During our study, we first tried to figure out the important factors in track\_popularity variable. For this purpose, after confirming linear regression assumptions, we used multiple regression analysis to find the linear correlation between Track Popularity and other variables. Then, after confirming the normality assumption for valance and track\_popularity variables, we used a two-sample t-test on the mean track popularity for different modality and a one sample t-test on the valence mean in order to decide on whether The Cure band's upbeat and euphoric songs are more popular than sad songs. For more analysis we also constructed a simple linear regression model for valance and popularity while also checking the regression assumptions. After this, we analyzed the intensity (the energy variable) for the 3 most popular albums of The Cure band namely 'Wish', 'Disintegration', and '4:13 Dream'. For this purpose, after verifying the required assumptions, we used a one-way ANOVA test. Then for the answer of question 4 we analyzed the instrumentalness of the same three albums. Unfortunately, this time the

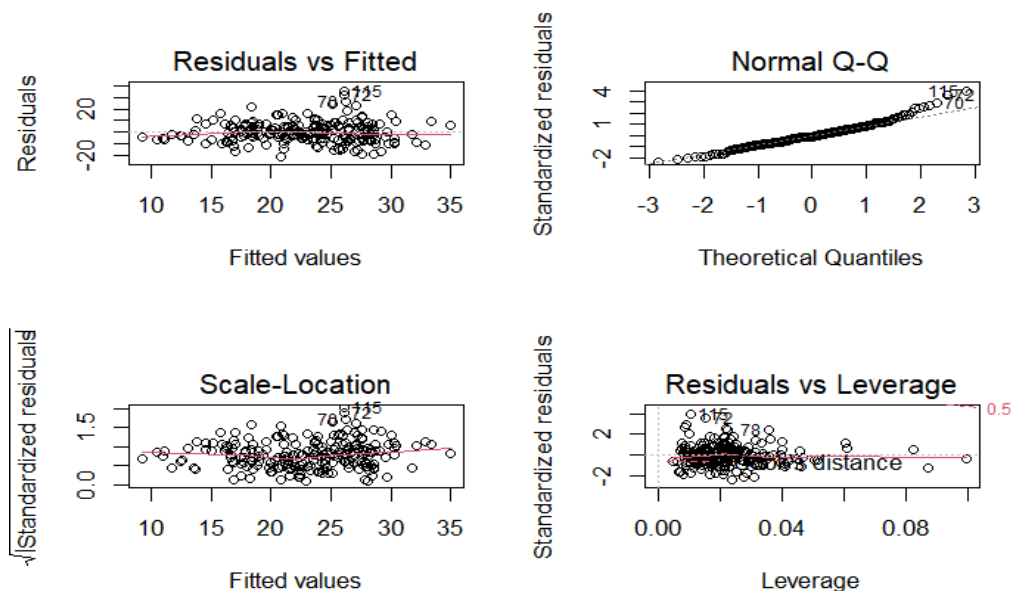
assumption of normality was rejected. So, we had to use a non-parametric equivalent test of one-way ANOVA namely, the Kruskal–Wallis’s test. After this, in order to check whether or not a high loudness effects the popularity of The Cure band’s style of rock music, we used a one-proportion z-test while also making sure that the required assumptions are true. And finally, in order to see whether high tempo (music speed) has a higher popularity on audience then a low tempo for The Cure’s songs, we checked the required assumptions and then used a two-proportion z-test.

### 3. Results and Findings

#### 3.1. What factors are correlated with Track-Popularity?

The variable of interest used in this analysis is Track Popularity. Multiple Regression Analysis has been used to find the linear correlation between Track Popularity and all other variables. The backward elimination method is used to find suitable predictor variables. Furthermore, linear regression assumptions have been diagnosed to check the eligibility of our model.

The result is as follows:



Based on the residuals graph, the mean of residuals has been distributed equally from each other, and the red line is approximately horizontal across all graphs. Moreover, the Normal Q-Q plot reveals that data residuals form a one-to-one line with theoretical residuals. Thus, the homoscedasticity and linearity assumptions are met. The table below shows the Shapiro normality test for the dependent variable

```

Shapiro-Wilk normality test
data: Track_Popularity
W = 0.99038, p-value = 0.1567

```

The P-Value is above .05 significant level; Thus, normality holds for the outcome variable.

Since regression analysis assumptions have been met, Multiple Regression analysis is appropriate. The table below shows the summary of regression analysis.

```

Call:
lm(formula = track_popularity ~ valence + danceability + liveness +
    loudness, data = cure_discography)

Residuals:
    Min     1Q  Median     3Q    Max
-20.830  -5.467  -1.185   4.715  35.893

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  27.7044    2.9858   9.279  < 2e-16 ***
valence       7.1359    3.4624   2.061  0.040489 *
danceability -14.2017    5.3578  -2.651  0.008622 **
liveness     -14.5003    1.9962  -7.264  6.59e-12 ***
loudness      -0.5765    0.1628  -3.541  0.000487 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.922 on 218 degrees of freedom
Multiple R-squared:  0.2326, Adjusted R-squared:  0.2185
F-statistic: 16.52 on 4 and 218 DF, p-value: 7.771e-12

```

Danceability, liveness, and loudness variables are negatively associated, while valence is positively associated with Track-Popularity. Furthermore, 23 percent of the variability in the outcome variable is explained by predictor variables.

### 3.2. When was the cure band more successful? When using major scale music or minor scale music?

We know that the type of the music that is played during a song, plays a major effect on the view of audience on that music and the message received from it. This is known that major chords create a happier and brighter sound for listeners while minor chords usually create a sad and more solemn environment. Here we want to check on how the change in the chords being played effects on the view of audience on The Cure Band's music. Because of the style of the rock music usually used by The Cure Band, we predict that their audience should prefer listening to major scale music style which has a brighter sound then minor scale style. Let's now analyze our hypothesis by a statistical test:

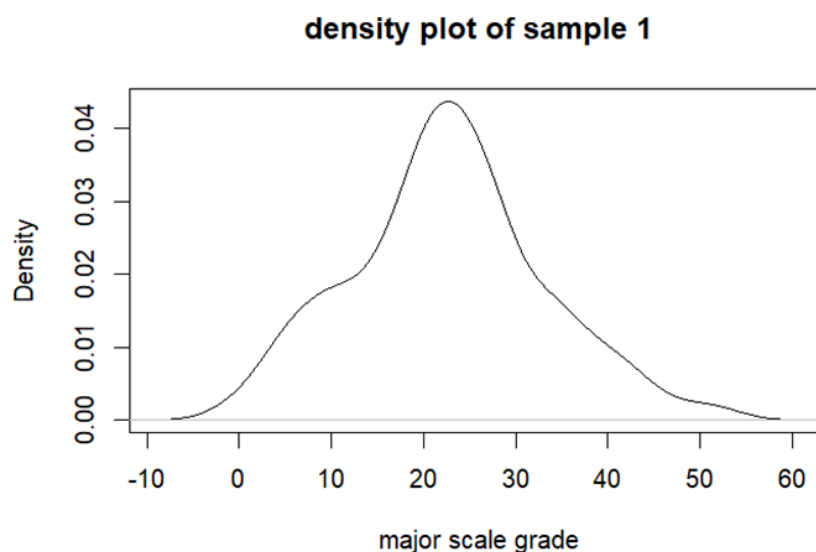
Here we have 2 samples. One for major chord popularity scores and one for minor chord popularity grades. The sample size for each sample is respectively 133 and 87. To check which sample has a higher mean freed, we decide on using a one-sided t-test of 2 samples as we know that the population variance is unknown here. First, we write our hypothesis and check the normality assumption for our 2 samples:

#### Our Hypothesis:

$$H_0: \mu_{\text{minor}} \geq \mu_{\text{major}}$$

$$H_A: \mu_{\text{minor}} < \mu_{\text{major}} \quad (\text{our claim})$$

Normality of sample one: lets first observe the density plot of our sample





It seems relatively normal. Let's make sure by using Shapiro test

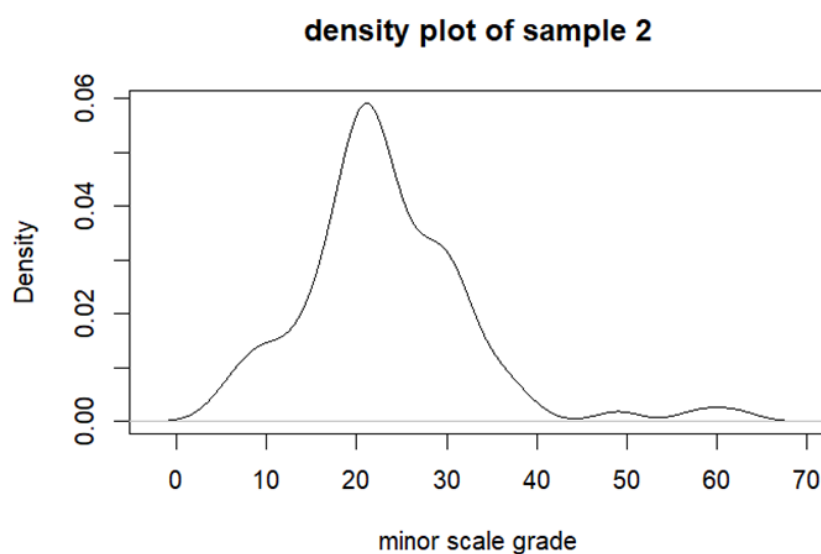
Shapiro-Wilk normality test

data: major

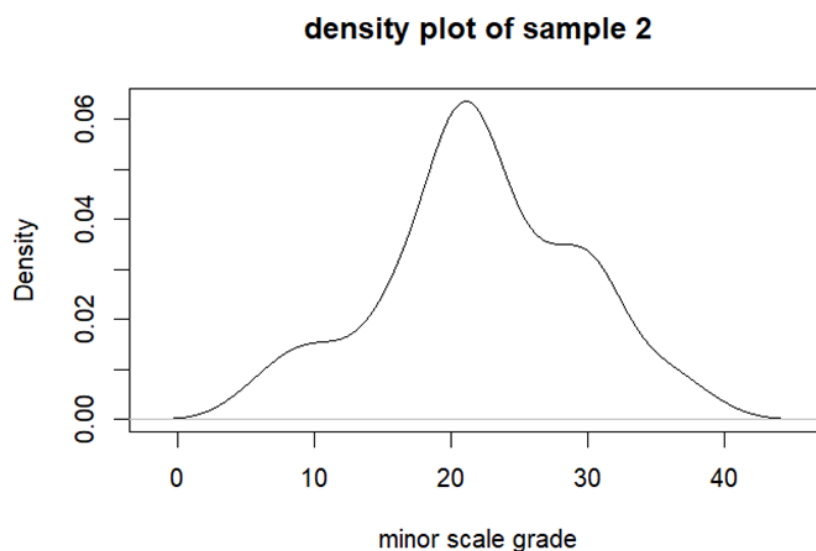
$W = 0.98773$ ,  $p\text{-value} = 0.2839$

Observe that the p-value is greater than 0.05. So, our first sample has a significant normality.

Now we will check the normality of the second sample. First observe the density plot:



It seems that we have some outliers. We will remove them and then decide on normality



It seems to have normality this time. We will make sure of this by a Shapiro test

Shapiro-Wilk normality test

data: minor

W = 0.98551, p-value = 0.4441

Observe that the p-value is greater than 0.05. So, our second sample also has a significant normality.

Now we can analyze our claim by doing a one-sided t-test on our 2 samples

Welch Two Sample t-test

data: major and minor

t = 0.38468, df = 216.97, p-value = 0.3504

alternative hypothesis: true difference in means is greater than 0

95 percent confidence interval:

-1.522834      Inf

sample estimates:

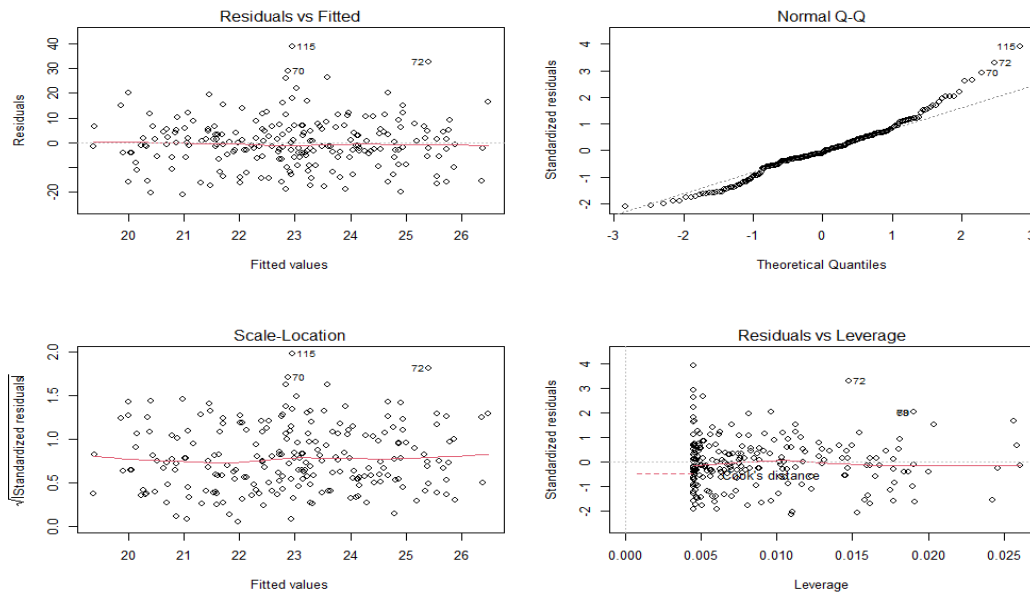
mean of x mean of y

22.66917 22.20690

We can observe that the t-test's p-value is equal to 0.3504 which is greater than 0.05. So, we can conclude that based on our data, we should reject null hypothesis. So here our claim could not be rejected and we conclude that major chord style is more popular with the style of The Cure band's music.

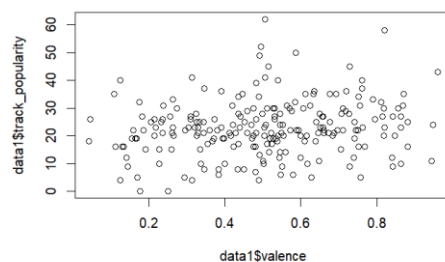
Now, since we know the change in happiness or sadness of a song for listeners has a direct relation with the modality of that song, then for doing extra analysis for question 2, we can also analyze the valence variable.

First, we do a simple linear regression for the valence and the track popularity.



By checking the above plots, we can check the simple linear regression assumptions:

- 1) by the method that our dataset was collected we know that our errors are independent.
- 2) By “Residual vs. Fitted”-plot we can conclude homoscedasticity
- 3) On Q-Q plot our points are mostly on a line → we have normality
- 4) Linearity between x and y: by the below scatter plot we can indeed observe that the points are more focused around a strain line:



So we create a model:

Call:

```
lm(formula = data1$track_popularity ~ data1$valence)
```

Residuals:

Min	1Q	Median	3Q	Max
-20.985	-5.466	-0.879	5.374	39.045

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	19.039	1.735	10.972	<2e-16 ***
data1\$valence	7.755	3.182	2.437	0.0156 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.982 on 221 degrees of freedom

Multiple R-squared: 0.02618, Adjusted R-squared: 0.02177

F-statistic: 5.941 on 1 and 221 DF, p-value: 0.01558

Both of our p-values are greater than 0.05 and our slope value is estimated to be 7.755. So, by a 95% confidence we can reject the null hypothesis that  $\beta_0$  and  $\beta_1$  are zero and we can see that there is a relatively strong positive linear relation between the valence and the popularity of each song for The Cure band.

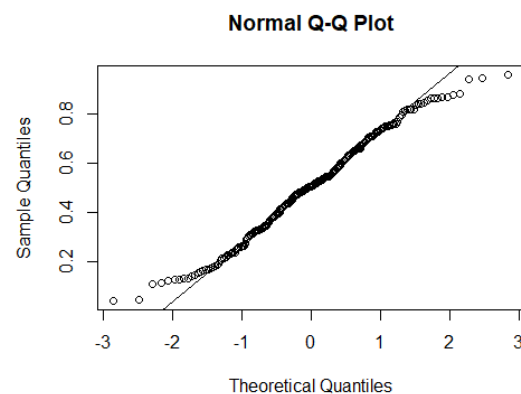
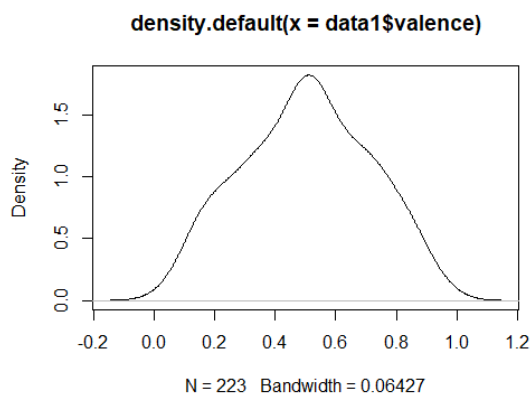
By the above simple linear regression, we saw that valence is positively associated with song popularity. Now we also formulate a two-sample hypothesis test to check whether The Cure Music band's upbeat and euphoric songs are more popular than sad songs. The result is as follows: (the valence higher than 0.45 is considered a positive song)

### Our Hypothesis:

$$H_0: \mu_{\text{valence}} \leq 0.45$$

$$H_A: \mu_{\text{valence}} > 0.45 \quad (\text{our claim})$$

Test of normality:



Again, we can see that our sample is approximately normal. Now we can use t-test:

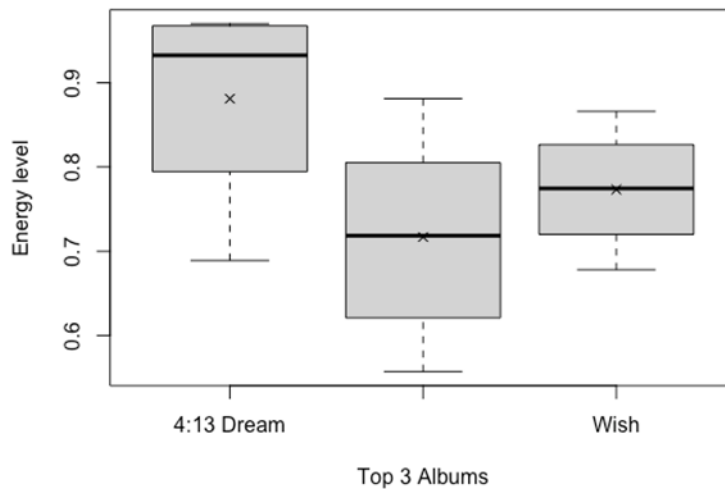
```
One Sample t-test

data: data1$valence
t = 3.7779, df = 222, p-value = 0.0001016
alternative hypothesis: true mean is greater than 0.45
95 percent confidence interval:
 0.4799816      Inf
sample estimates:
mean of x
0.5032731
```

p-value<0.05 so by a 95% confidence we reject the null hypothesis here. So, we conclude that here, our claim is correct. Thus, based on the analysis above, the Cure Music band's euphoric songs were more popular than their downbeat songs.

### **3.3. Is there any statistically significant difference between the mean energy levels of the albums in the three most popular albums at 0.05 significance level?**

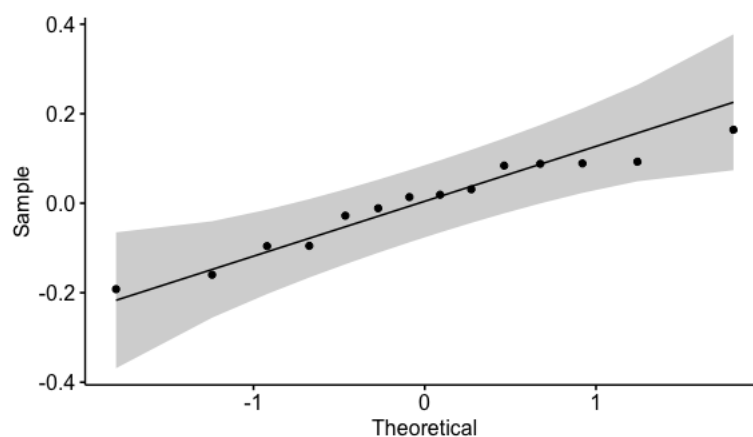
Music plays an important role one's life. We are surrounded by music wherever we go. Sometimes it might be pleasant for someone and sometimes not. Recently, people have started to listen to more upbeat and energetic music which makes one dance. People have started to associate themselves much more with high energy songs and the days of slow, classical music are long gone. Here we have tried to analyze whether this phenomenon has been true lately. First of all, we have filtered out the most popular albums from our dataset and selected the top 3 of them which are Wish, Disintegration (Deluxe Edition [Remastered]) and 4:13 Dream. In order to analyze the mean energy levels of these 3 albums, we must satisfy assumptions for One-way ANOVA first.



Here the crosses on the boxplot show the means of energy level of each album. Two of them seem to be close to each other but we must check if there is statistical difference between the 3 of them or not?

Checking the assumptions now:

- 1) Independence: The samples for each album are independent as there is no relationship between observations in each group.
- 2) Checking for outliers: Using identify outliers function in rstatix library we come to the conclusion that there are no extreme outliers for the variable we are investigating.
- 3) Normality:



**QQ Plot for energy and Top 3 Albums**

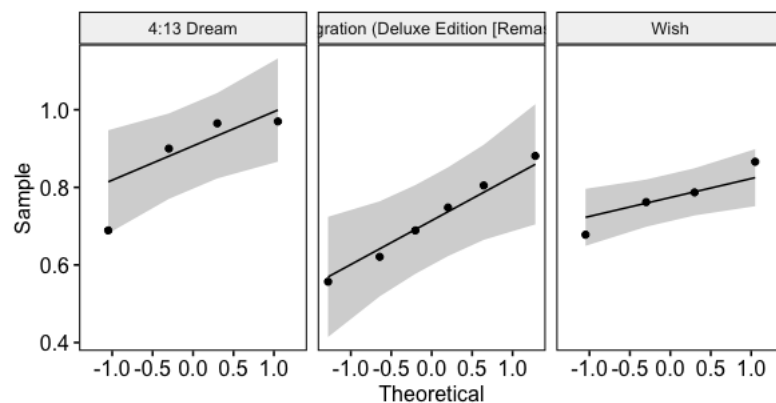
Shapiro-Wilk Normality Test:

Data: Energy levels for Top 3 Albums

$W = 0.950$  ,  $p\text{-value} = 0.554$

Using the Q-Q plot and Shapiro test which gives the  $p\text{-value} = 0.554$  we conclude that these variables are normal. Since the  $p\text{-value}$  is greater than 0.05 for Shapiro test and most of points are close to the best fit line in QQ-plot we can assume normality. Therefore, we can use One-way ANOVA for this case.

- 1) Normality by groups: Again, Shapiro test is applied for all 3 albums. The  $p\text{-values}$  for each of them is greater than 0.05 which means all of them are normal. The following Q-Q plots for each album show us the same.



Shapiro-Wilk Normality Test:

Data: Energy levels for Top 3 Albums (For each album)

$W(\text{Wish}) = 0.989$

$p\text{-value}(\text{Wish}) = 0.950$

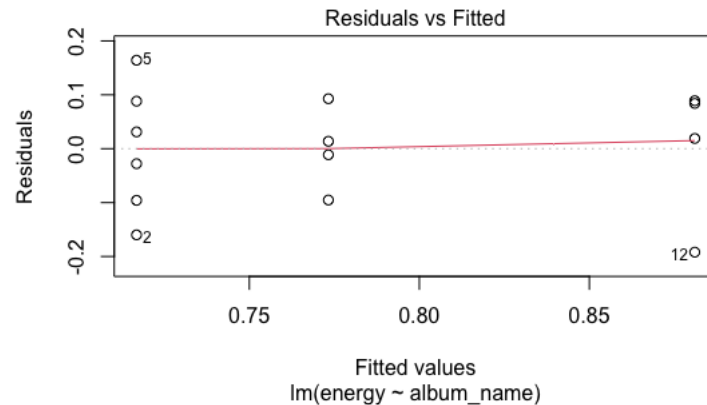
$W(4:13 \text{ Dream}) = 0.797$

$p\text{-value}(4:13 \text{ Dream}) = 0.0964$

$W(\text{Disintegration...}) = 0.988$

$p\text{-value}(\text{Disintegration...}) = 0.983$

## 2) Equality of Variances:



Apparently in the graph, there is no evidence of any relationship between the residuals and fitted values. So, we can assume equality of variances here. Moreover, make this finding much more solid we also got the ratio of the largest and smallest standard deviations. The value of the ratio was 1.703409 which is lesser than 2 and proves equality once more. After proving all the assumptions, we now state our hypotheses.

### Our Hypothesis:

$H_0: \mu_{\text{Wish}} = \mu_{\text{4:13 Dream}} = \mu_{\text{Disintegration (Deluxe Edition [Remastered])}}$

$H_A$ : The means for at least one energy level is not equal to others are not equal.

Thereafter we carry out the ANOVA test and reach to the following conclusions.

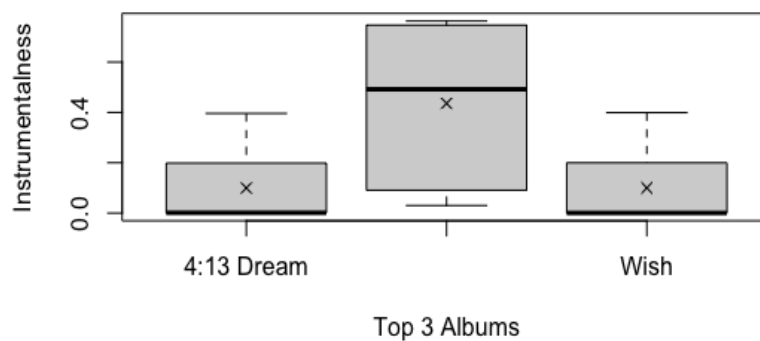
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
album_name	2	0.06493	0.03246	2.526	0.125
Residuals	11	0.14139	0.01285		

Since the p-value = 0.125 > 0.05 there's not enough evidence to reject the Null Hypothesis at 95% Confidence level. As a result, we come to the conclusion that the mean energy levels for 3 most popular albums are equal and there's no significant difference between them.



### 3.4. Can we claim that the instrumentality of the songs is also effective on how the track effects the audience for the tracks in question 3?

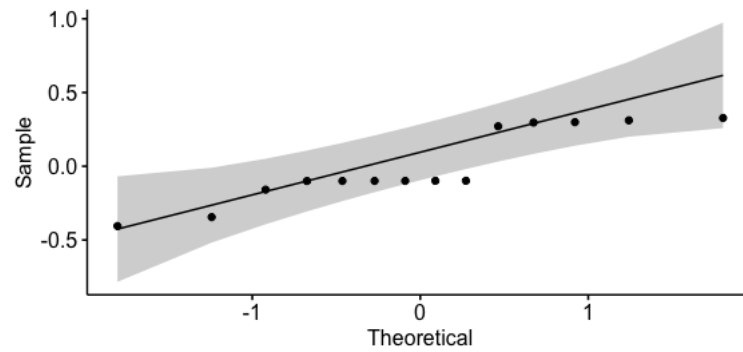
Building upon from the 2<sup>nd</sup> question, here we have tried to analyze whether instrumentality has been a factor that affects the audience and influences a songs popularity among them. Like the third question, firstly we have filtered out the most popular albums from our dataset and selected the top 3 of them which are Wish, Disintegration (Deluxe Edition [Remastered]) and 4:13 Dream. In order to analyze the mean instrumentality of these 3 albums, we must satisfy assumptions for One-way ANOVA first.



Here the crosses on the boxplot show the means of instrumentality of each album. None of them seem to be close to each other but we must check if there is statistical difference between the 3 of them or not using some other statistical procedures.

Checking the assumptions now:

- 1) Independence: The samples for each album are independent as there is no relationship between observations in each group.
- 2) Checking for outliers: Using identify outliers function in rstatix library we come to the conclusion that there are no extreme outliers for the variable we are investigating.
- 3) Normality:



### QQ Plot for Instrumentalness and Top 3 Albums

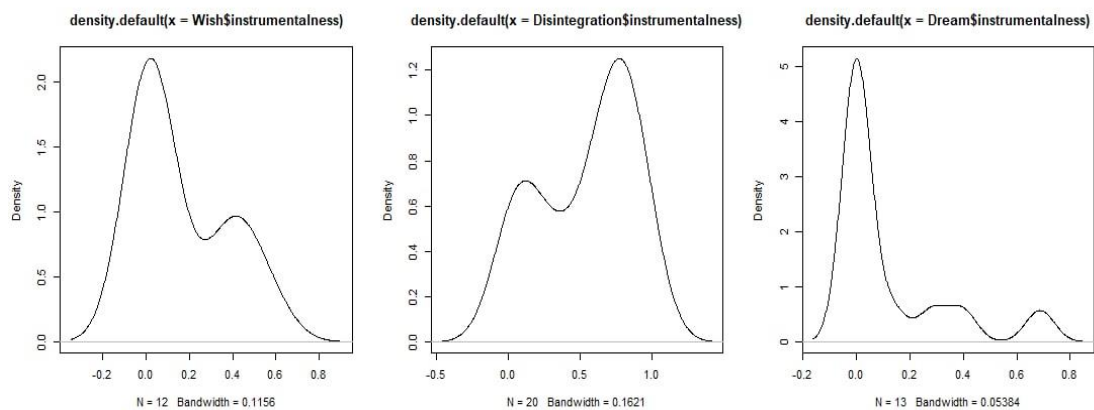
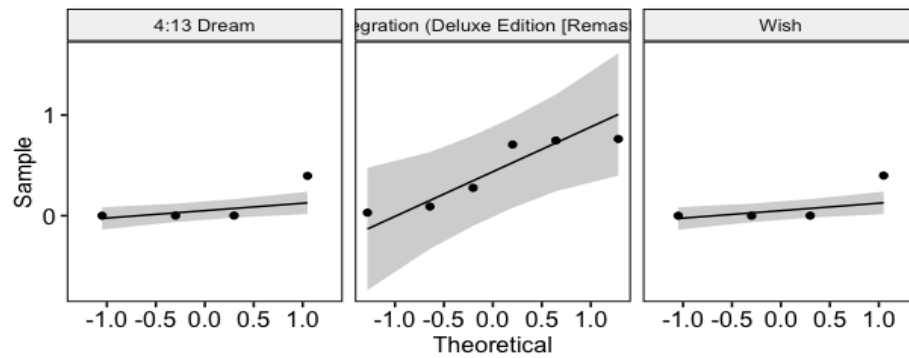
Shapiro-Wilk Normality Test:

Data: Instrumentalness for Top 3 Albums

$W = 0.843$ ,  $p\text{-value} = 0.0178$

Using the Q-Q plot and Shapiro test which gives the  $p\text{-value} = 0.0178$  we conclude that these variables are not normal. Since the  $p\text{-value}$  is lesser than 0.05 for Shapiro test and majority of points lie far away from the best fit line in QQ-plot, we cannot assume normality. Therefore, we fail to use the One-Way ANOVA. Instead of that we will use Kruskal Wallis Test now.

- a. Normality by groups: Again, Shapiro test is applied for all 3 albums. The  $p\text{-values}$  for each of them is lesser than 0.05 which means all of them are not normal. The following Q-Q plots (as most points are not on the best fit line) and density plots for each album show us the same (Not bell shaped).



### Shapiro-Wilk Normality Test:

Data: Instrumentalness for Top 3 Albums (For each album)

W(Wish) = 0.631

p-value(Wish) = 0.00129

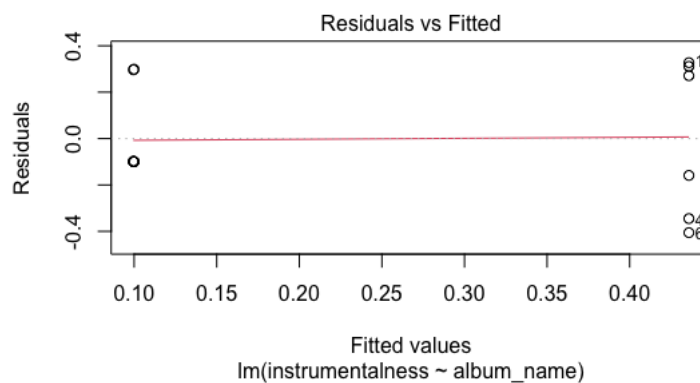
W(4:13 Dream) = 0.820

p-value(4:13 Dream) = 0.00883

W(Disintegration...) = 0.988

p-value(Disintegration...) = 0.983

### 1) Equality of Variances:



Apparently in the graph, there is no evidence of any relationship between the residuals and fitted values. So, we can assume equality of variances here. Moreover, we also got the ratio of the largest and smallest standard deviations. The value of the ratio was 1.732427 which is lesser than 2 and proves equality once. This shows there is equality in variances, but still, we cannot apply the ANOVA test since normality assumption doesn't hold. Therefore, we apply Kruskal Wallis Test.

Stating out hypotheses first.

### **Our Hypotheses**

$H_0$ : Median<sub>Wish</sub> = Median<sub>Disintegration (Deluxe Edition [Remastered])</sub> = Median<sub>4:13 Dream</sub>

$H_A$ : At least one median for instrumentalness is not equal to others.

Thereafter we carry out the Kruskal Wallis test and reach to the following conclusions.

Data: Instrumentalness for Top 3 Albums (For each album)

Kruskal Wallis chi squared = 5.6571

df = 2

p-value = 0.0591

Since the p-value = 0.0591 > 0.05 there's not enough evidence to reject the Null Hypothesis at 95% Confidence level. As a result, we come to the conclusion that the median instrumentalness for 3 most popular albums is equal and there's no significant difference between them.

### **3.5. In modern rock music, the famous bands usually create song which have a high loudness. Was this factor also important in The Cure band's music?**

By the use of loudness variable, we can see how loud each of The Cure tracks were at the time that they were played. We predict that more than 50 percent of The Cure band's popular songs have a high loudness value. Now we check our assumption more, by the use of one proportion z-test. (We know that loudness > -8.5 is considered a loud song. Also, a popularity grade higher than 20 is considered a popular song)

### Our Hypothesis:

$$H_0: P_{\text{high loudness tracks}} \leq 0.5$$

$$H_A: P_{\text{high loudness tracks}} > 0.5 \quad (\text{our claim})$$

First, we check if  $np > 5$  and  $n(1-p) > 5$  conditions are met so that we can assume a normality distribution for our test:  $n = 71$ ,  $\hat{p}_{\text{high loudness tracks}} = 0.522$

So, we have  $np = 37.06$  and  $n(1-p) = 33.93$ . This shows that our sample is big enough to assume normality.

Now we can analyze our hypothesis by a one-sided one-proportion z-test:

1-sample proportions test with continuity correction

data: 71 out of 136, null probability 0.5

X-squared = 0.18382, df = 1, p-value = 0.3341

alternative hypothesis: true p is greater than 0.5

95 percent confidence interval:

0.4482387 1.0000000

sample estimates:

p

0.5220588

We can see that  $p\text{-value} > 0.05$ . So, by a 95% confidence we can reject the null hypothesis. This shows us that our claim is correct and more than half of The Cure band's popular song's have a high loudness value.

### **3.6. It is usually considered by people that, in the rock genre more popular tracks are played at a high speed. Is this also a correct for The Cure?**

Tempo variable in our data shows us the tracks speed. And we consider the tracks with a track\_popularity higher then 20 as popular tracks. We also know that a tempo greater than 120 shows us a fast speed for each track. Now we expect that the proportion of popular songs with high tempo is higher than the sings with low tempo. We analyze our hypothesis by a z-test for 2 proportions:

### Our Hypothesis:

$$H_0: P_{\text{tempo high}} \leq P_{\text{tempo low}}$$

$$H_A: P_{\text{tempo high}} > P_{\text{tempo low}} \quad (\text{our claim})$$

Assumptions of z-test for 2 proportions:

- The samples are randomly selected. → True based on how our data was collected
- The samples are independent. → True. Since one sample shows the number of songs with a high tempo while the other is about the songs with a low tempo.

c.  $n_1 \hat{p}_1 > 5$  and  $n_2 \hat{p}_2 > 5$  and  $n_1(1-\hat{p}_1) > 5$  and  $n_2(1-\hat{p}_2) > 5$ :

$$n_1 = 75, n_2 = 61, \hat{p}_1 = 0.55, \hat{p}_2 = 0.44$$

$$n_1 \hat{p}_1 = 41.36 \text{ and } n_2 \hat{p}_2 = 27.36 \text{ and } n_1(1-\hat{p}_1) = 33.63 \text{ and } n_2(1-\hat{p}_2) = 33.63$$

So, the samples are large enough to use a normal sampling distribution.

Now we can do a 2-proportion z-test:

2-sample test for equality of proportions with continuity correction

data: c(75, 61) out of c(136, 136)

X-squared = 2.4853, df = 1, p-value = 0.05746

alternative hypothesis: greater

95 percent confidence interval:

-0.003615824 1.000000000

sample estimates:

prop 1 prop 2

0.5514706 0.4485294

We can see here that  $p\text{-value} > 0.05$  so by a 95% confidence we conclude that we are not able to reject the null hypothesis. So, our claim is not correct and we can see that unlike what many people might think, not all the famous songs of The Cure had very high tempo.

#### **4. Discussion/Conclusion**

In conclusion, based on the analysis, the Cure rock band's euphoriant songs were well perceived. Contrarily, their downbeat songs were not that popular. The reason could be that most rock fans prefer upbeat songs to downbeat ones. Moreover, the analysis showed that the danceability of the famous songs was inversely correlated with song popularity. The rationale might be due to chance, or it could be that another variable governs both danceability and track popularity.

Furthermore, the Cure Rock band's famous music mainly exhibited major scales such as higher chords. This is true for every rock music since major scale music is an attribute to rock music. Thus, the Cure rock band did not showcase creativity in that regard. An interesting finding was that the Cure rock band's top three music albums were energetic. It shows the amount of creativity and showmanship they put in their music to exhibit such a dynamic range of activity. This could be an attribute that would distinguish them from a typical rock band. Also, our findings are in tandem with the general idea that modern rock music bands typically create songs with a high loudness threshold, and thus, the Cure rock band also exhibits similar traits. Most rock fans assume that popular tracks are played at high speed; on the contrary, based on our findings, the cure rock band's well-known songs did not showcase such traits.

## ***References***

*The Cure* discography. (2018, August 12). Kaggle.

<https://www.kaggle.com/datasets/xvivancos/the-cure-discography>

*Linear Regression Assumptions and Diagnostics in R: Essentials*. (2018, March 11). Articles -

STHDA. <http://www.sthda.com/english/articles/39-regression-model-diagnostics/161-linear-regression-assumptions-and-diagnostics-in-r-essentials/>

*One-Proportion Z-Test in R - Easy Guides - Wiki - STHDA*. (n.d.). Sthda.

<http://www.sthda.com/english/wiki/one-proportion-z-test-in-r>

*Two-Proportions Z-Test in R - Easy Guides - Wiki - STHDA*. (n.d.). Sthda.

<http://www.sthda.com/english/wiki/two-proportions-z-test-in-r>

W. (2022, June 28). *The Role of Music & Songs in Human Communication*. WhisperRoom,

Inc.<sup>TM</sup>. <https://whisperroom.com/tips/what-is-acoustics/#:%7E:text=Acoustics%20is%20a%20term%20referring,control%2C%20transmission%2C%20and%20reception.>