MUSIC LISTENING HABITS AND ACADEMIC PERFORMANCE

To what extent does a person's music listening habits predict their academic performance in a core subject area?

Sumantra Chattopadhyay

Mrs. Hudak

AP Statistics Block: 1

13 June 2018

Table of Contents

Introduction

Research Question

Through my research, I aim to answer the question—to what extent does a person's music listening habits predict their academic performance in a core subject area? This research question uses certain terminology that are defined below:

- *Music listening habits*- includes (but may not be limited to) the following:
    - the genre of music a person listens to on average
    - the number of hours a person spends listening to music on average
- *Academic performance*- reflection of a person's understanding of the subject matter and its applications; measured by marking period grades earned in a particular subject
- *Core subject area*- mathematics, English, history and science are considered core subject areas

Variables

The main variables used in this project are music preference and academic performance.

Music preference is a categorical variable because each survey respondent selects a particular genre from an extensive list of given options. I chose this variable because it is the explanatory variable of my research question.

In contrast, academic performance is a discrete quantitative variable because each survey respondent gives a positive integer answer. I chose this variable because it is the response variable of my research question. Specifically, academic performance is measured by asking each respondent's marking period 3 average. It is assumed that they will have responded with the unweighted average and I will compute the weighted average based on the level of the classes they are taking. Additionally, academic performance will be measured by calculating

the unweighted and weighted averages of the marking period 3 averages of the 4 core subject areas[1].

Population and Sampling

The targeted population of this research is all high school students. Specifically, the population is all of the students at Morris Knolls High School that are currently enrolled in at least one mathematics class because this is where I sampled from.

I used a simple random sample (SRS) method to choose participants who would receive the survey. First, a TI-84 Plus graphing calculator was set to a certain "seed" for its random number generation feature. Then, 40 random numbers were generated using the calculator. Each student in the population already had a number assigned to them by alphabetical order. The 40 non-repeating random numbers were matched with 40 students from the population to create a sample of $n = 40$ to receive the survey. However, due to certain erroneous answers provided by some participants, in the end, there were 30 usable surveys. Thus, in the end $n = 30$.

Prediction

I predict that a Morris Knolls High School student's music listening habits will have a statistically significant impact on their academic performance in core subject areas. Thus, music listening habits will, to a large extent, predict academic performance in a core subject area.

[1] If a respondent is not taking 4 core subject classes, then the unweighted and weighted averages of however many core classes they are taking will be calculated.

Analysis of Numerical Variables

Class Levels

CPB – How many CPB classes are you taking this year?

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
CPB Classes	30	0.7333	0.2667	1.4606	2.1333	0.0000	0.0000	0.0000	0.2500	4.0000	4.0000	0.2500	0

Figure 1.1: Distribution of the Number of CPB Classes Being Taken, Boxplot

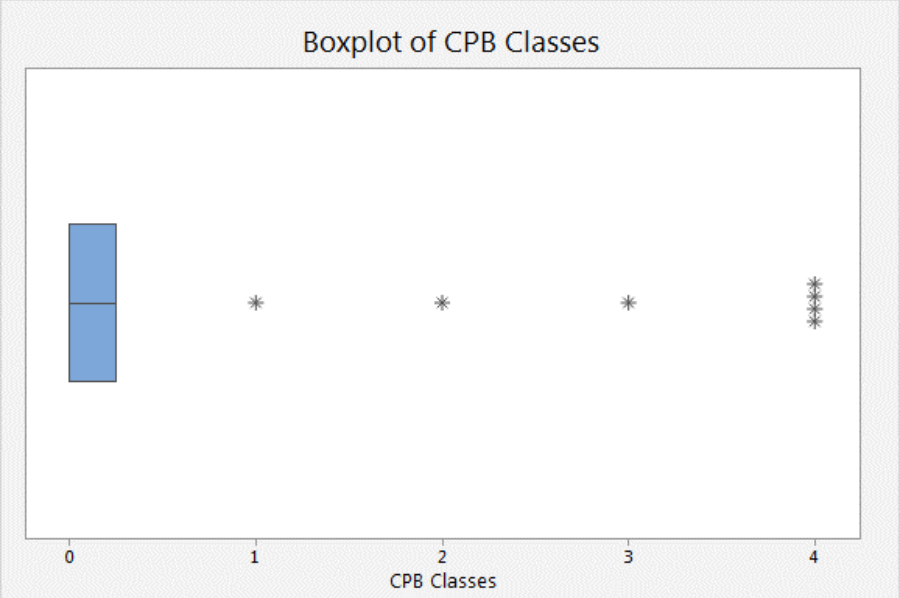
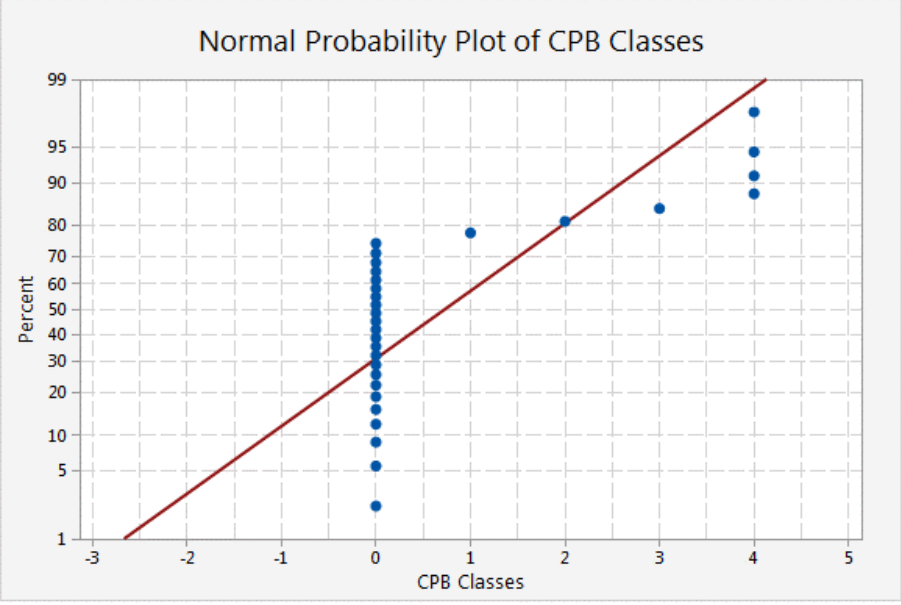


Figure 1.2: Distribution of the Number of CPB Classes Being Taken, NPP



These graphs show that the distribution of the number of CPB classes being taken by the survey respondents have 7 outliers—1 is taking 1 CPB class, 1 is taking 2 CPB classes, 1 is taking 3 CPB classes, and 4 are taking 4 CPB classes. Thus, 23 respondents are taking 0 CPB classes. This results in the median being 0 CPB classes but the mean being 0.7333 CPB classes. Clearly, the right-skewness of the distribution and the mean being greater than the median should result in the median of 0 CPB classes being the center of this distribution. The 23 non-outlier data points are clustered at 0 CPB classes while the other 7 data points are spread further from this center. Minitab Express calculated the standard deviation of these data to be 1.4606, which is a moderate spread. The NPP clearly shows a non-linear trend, indicating that the data are not normally distributed.

CPA – How many CPA classes are you taking this year?

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
CPA Classes	30	2.3000	0.3900	2.1359	0.0000	0.0000	2.0000	4.0000	6.0000	6.0000	4.0000	0

Figure 2.1: Distribution of the Number of CPA Classes Being Taken, Boxplot

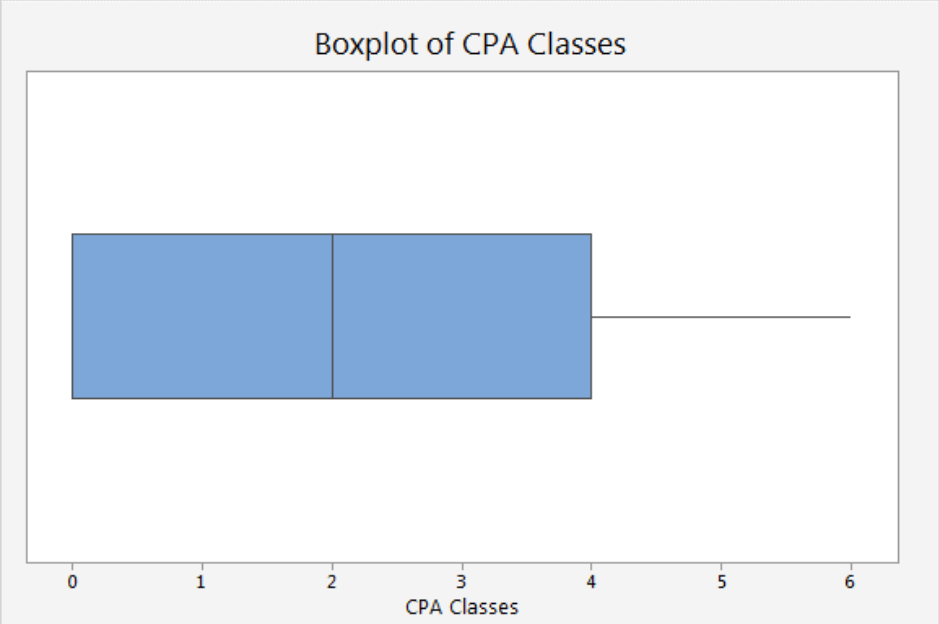
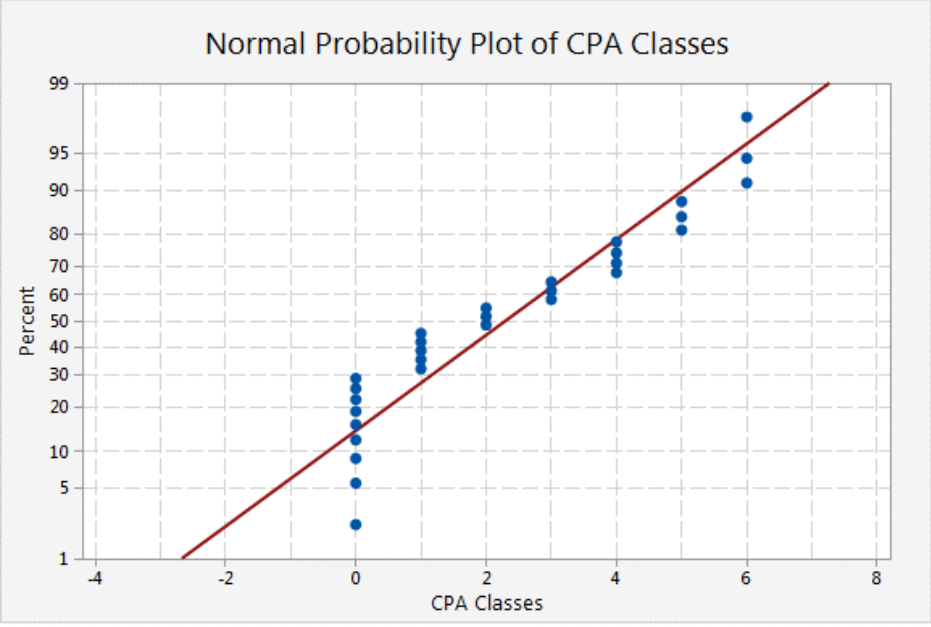


Figure 2.2: Distribution of the Number of CPA Classes Being Taken, NPP



These graphs show that the distribution of the number of CPA classes being taken by the survey respondents have 0 outliers. This results in the median being 2 CPA classes and the mean being 2.3000 CPA classes. Clearly, the graphs show moderate right-skewness of the distribution and the mean being slightly greater than the median should result in the median of 2 CPA classes being the center of this distribution. There is greater spread in this distribution as Minitab Express calculated the standard deviation of these data to be 2.1359. The NPP shows a slight linear trend, but the data do not appear to be normally distributed.

Honors – How many Honors classes are you taking this year?

Variable	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
Honors Classes	30	0.9667	0.2690	1.4735	0.0000	0.0000	0.0000	2.0000	5.0000	5.0000	2.0000	0

Figure 3.1: Distribution of the Number of Honors Classes Being Taken, Boxplot

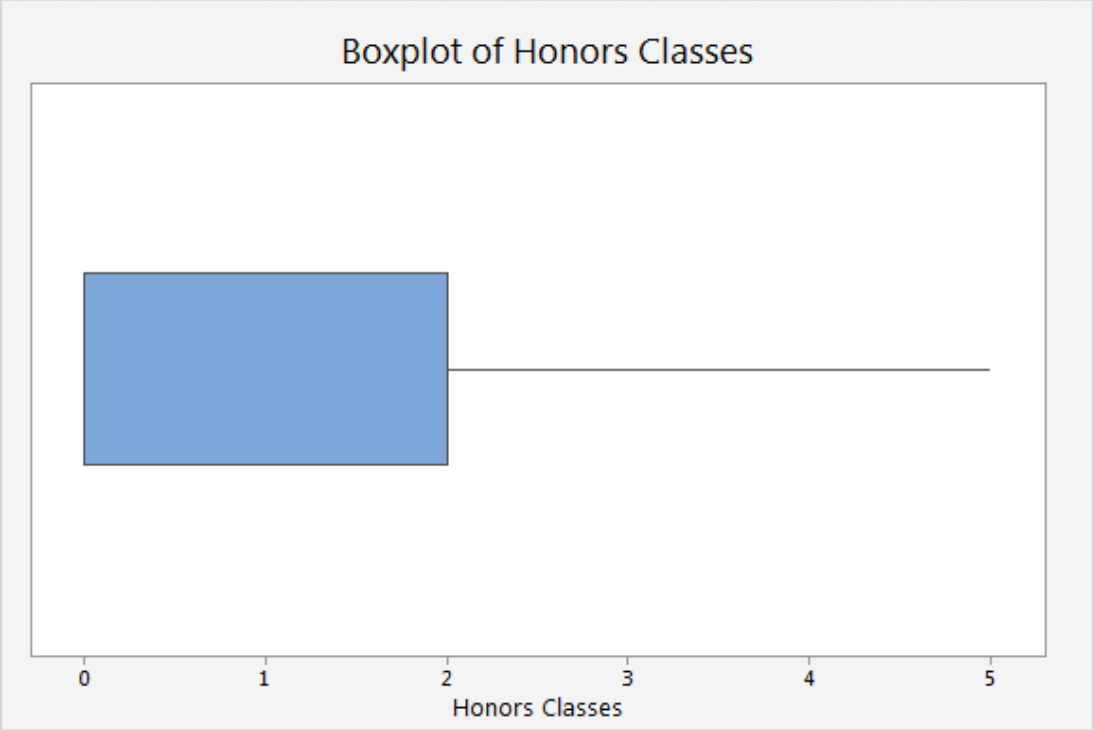
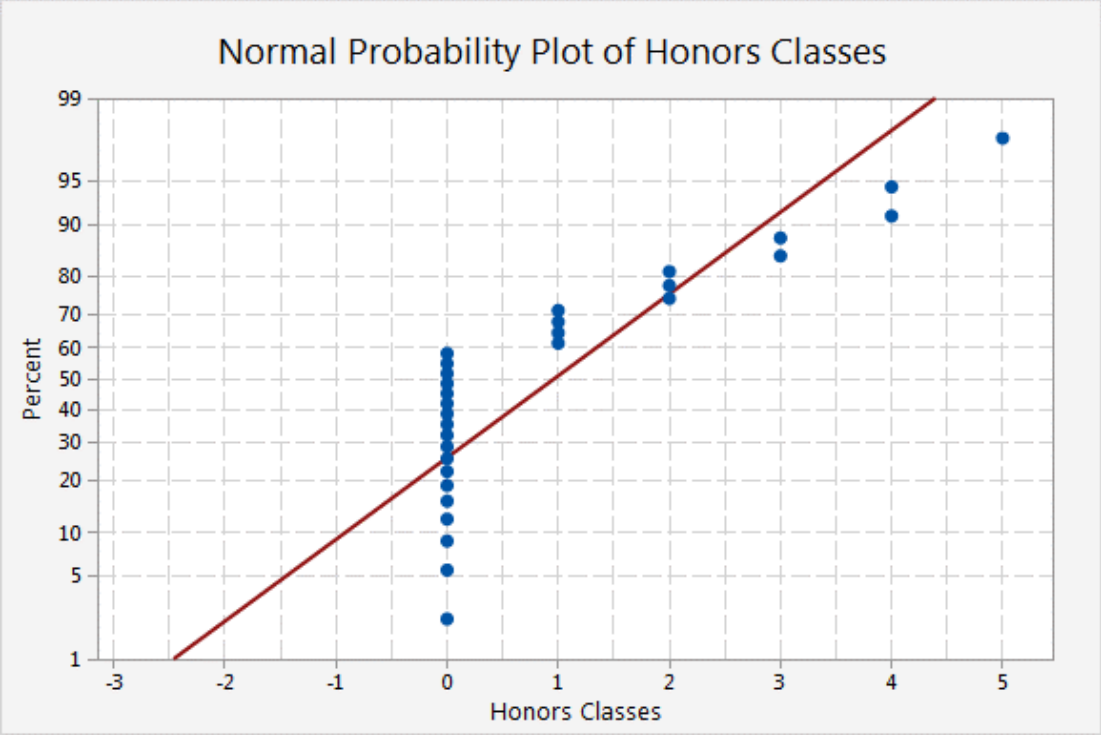


Figure 3.2: Distribution of the Number of Honors Classes Being Taken, NPP



These graphs show that the distribution of the number of Honors classes being taken by the survey respondents have 0 outliers. This results in the median being 0 Honors classes and the mean being 0.9667 Honors classes. Clearly, the graphs show strong right-skewness of the distribution and the mean being slightly greater than the median should result in the median of 0 Honors classes being the center of this distribution. There is moderate spread in this distribution as Minitab Express calculated the standard deviation of these data to be 1.4735. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

AP/IB – How many AP/IB classes are you taking this year?

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
AP/IB Classes	30	1.4000	0.4113	2.2530	5.0759	0.0000	0.0000	0.0000	2.0000	8.0000	8.0000	2.0000	0

Figure 4.1: Distribution of the Number of AP/IB Classes Being Taken, Boxplot

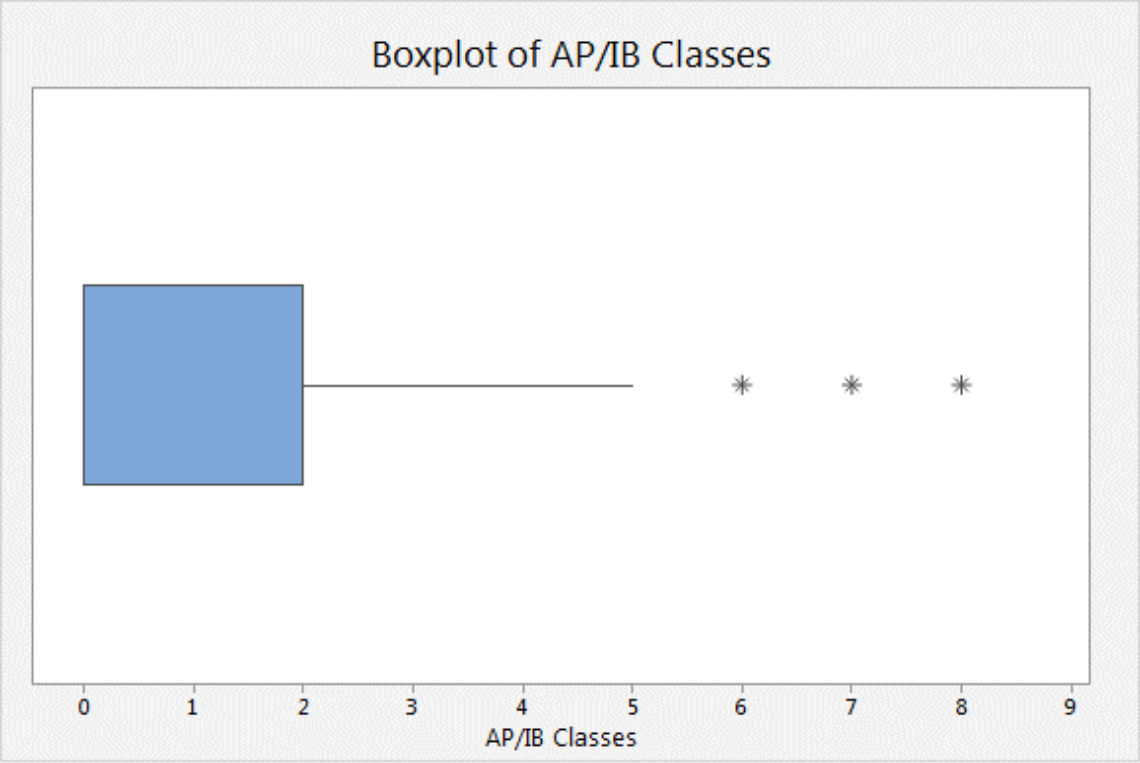
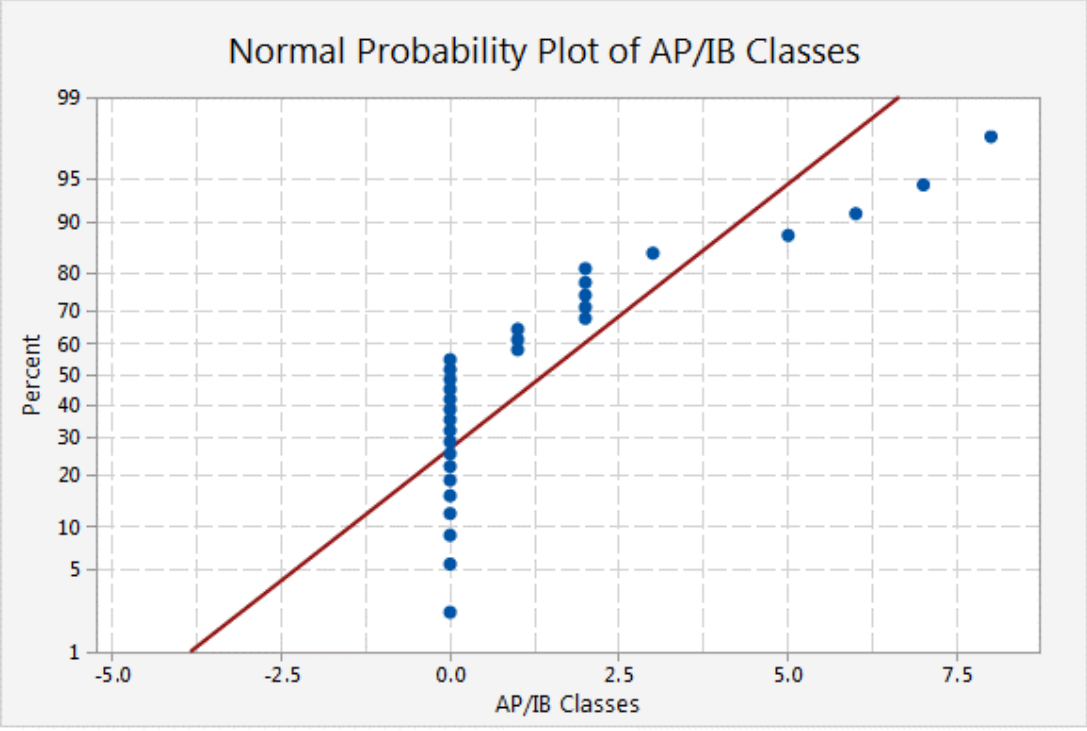


Figure 4.2: Distribution of the Number of AP/IB Classes Being Taken, NPP



These graphs show that the distribution of the number of AP/IB classes being taken by the survey respondents have 3 outliers—1 person is taking 6 AP/IB classes, 1 person is taking 7 AP/IB classes, and 1 person is taking 8 AP/IB classes. This results in the median being 0 AP/IB classes and the mean being 1.4000 AP/IB classes. Clearly, the graphs show strong right-skewness of the distribution and the mean being moderately greater than the median should result in the median of 0 AP/IB classes being the center of this distribution. There is greater spread in this distribution as Minitab Express calculated the standard deviation of these data to be 2.2530. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

Class Averages

History – What was your Marking Period 3 average in history?

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
History MP3 Avg (unweighted)	28	90.2143	0.9076	4.8024	23.0635	79.0000	87.2500	90.0000	93.0000	99.0000	20.0000	5.7500	89, 90, 93

Figure 1.1: Distribution of MP3 History Unweighted Averages, Boxplot

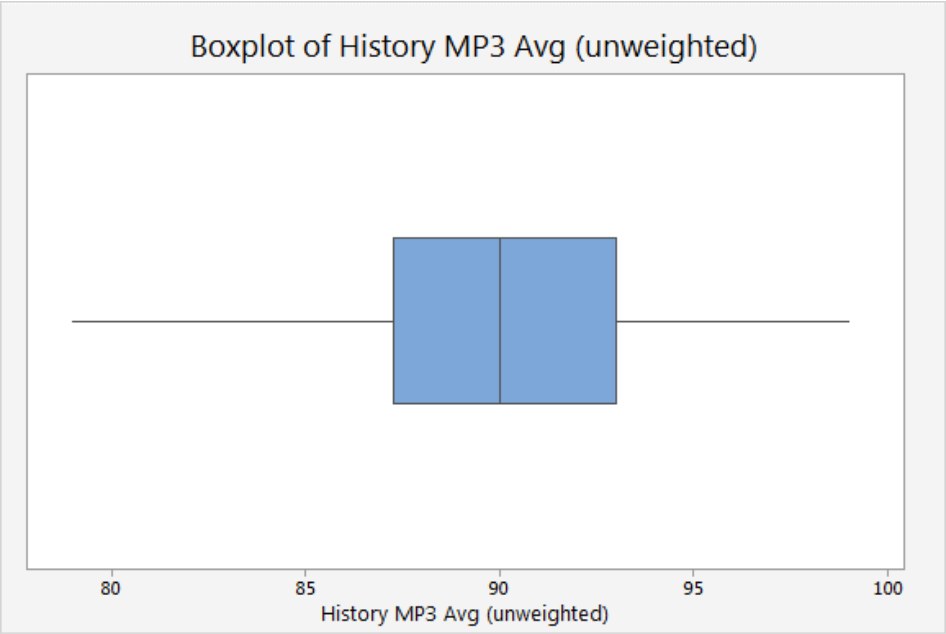
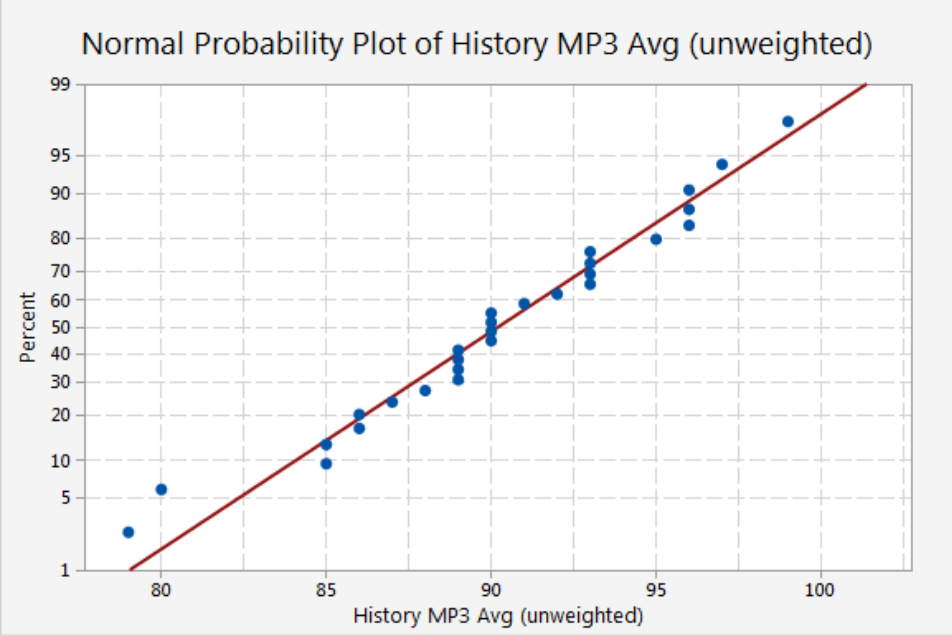


Figure 1.2: Distribution of MP3 History Unweighted Averages, NPP



These graphs show that the distribution of the unweighted marking period 3 average in history, as reported by the survey respondents, have 0 outliers. This results in the median being 90 and the mean being 90.2143. Clearly, the graphs show approximate symmetry of the distribution and the mean being very slightly greater than the median can result in the mean of 90.2143 being the center of this distribution. There is moderate spread in this distribution as Minitab Express calculated the standard deviation of these data to be 4.8024. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

English – What was your Marking Period 3 average in English?

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
English MP3 Avg (unweighted)	30	90.233	1.242	6.801	46.254	70.000	85.750	91.000	95.000	102.000	32.000	9.250	85, 90, 93, 95

Figure 2.1: Distribution of MP3 English Unweighted Averages, Boxplot

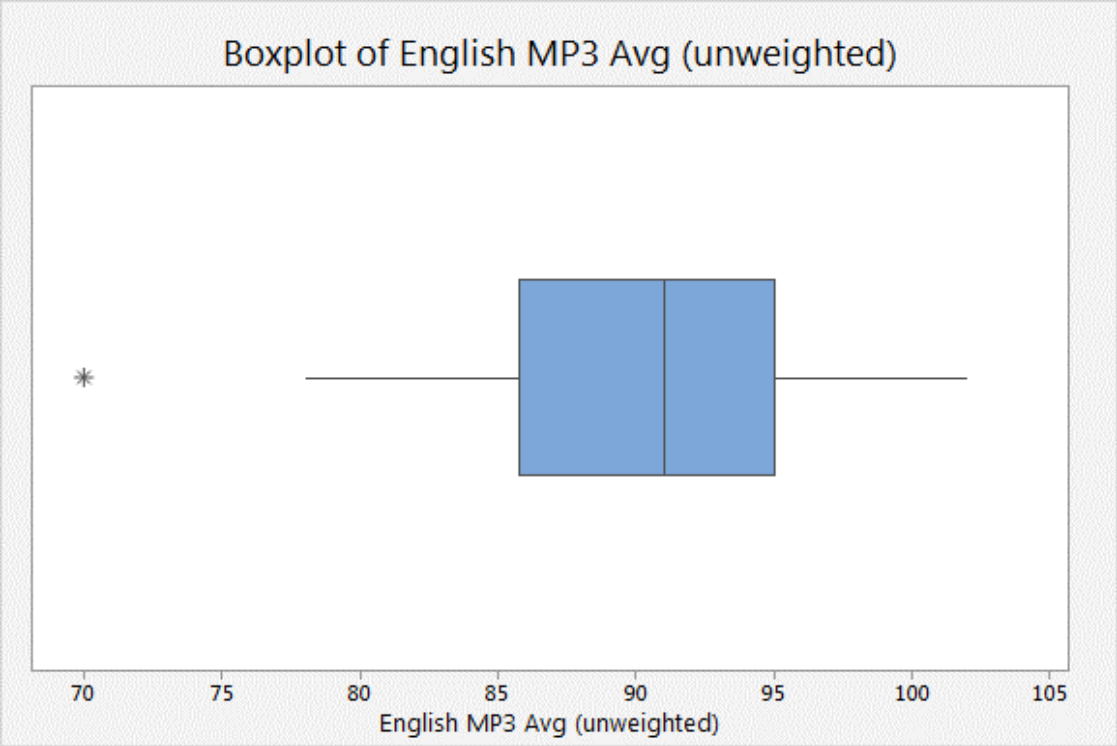
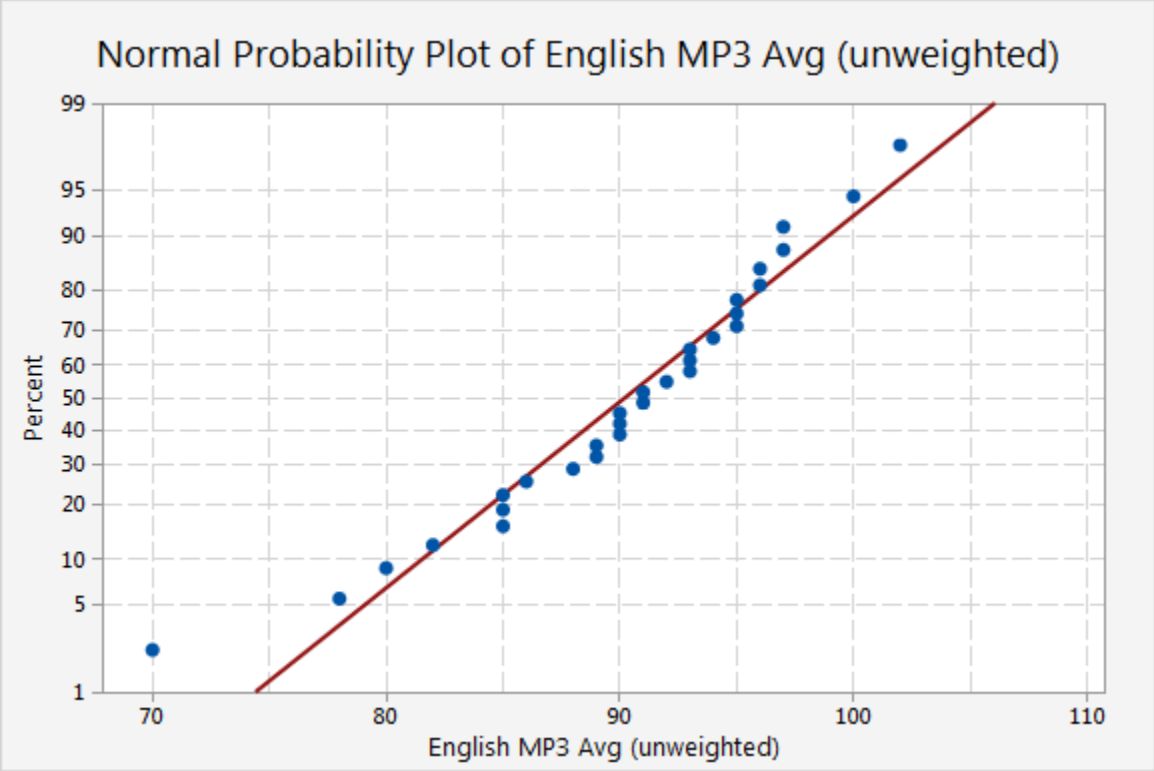


Figure 2.2: Distribution of MP3 English Unweighted Averages, NPP



These graphs show that the distribution of the unweighted marking period 3 average in English, as reported by the survey respondents, have 1 outlier of 70. This results in the median being 91 and the mean being 90.233. Clearly, the graphs show slight left-skewness of the distribution and the mean being slightly lower than the median should result in the median of 91 being the center of this distribution. There is greater spread in this distribution as Minitab Express calculated the standard deviation of these data to be 6.801. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

Math – What was your Marking Period 3 average in Math?

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
Math MP3 Avg (unweighted)	30	88.000	1.534	8.400	70.552	69.000	81.250	90.500	93.250	100.000	31.000	12.000	91

Figure 3.1: Distribution of MP3 Math Unweighted Averages, Boxplot

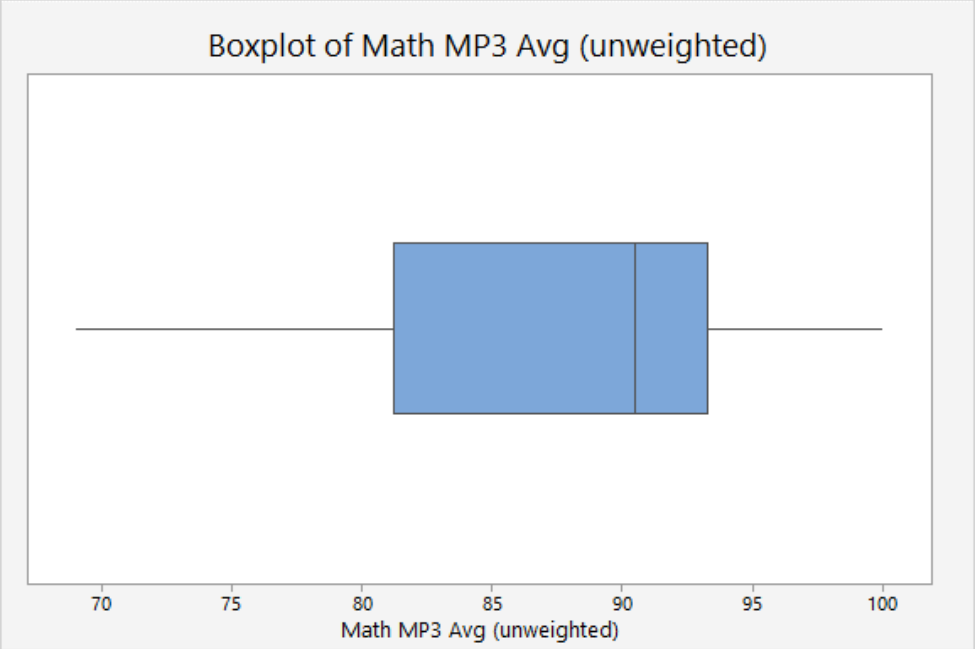
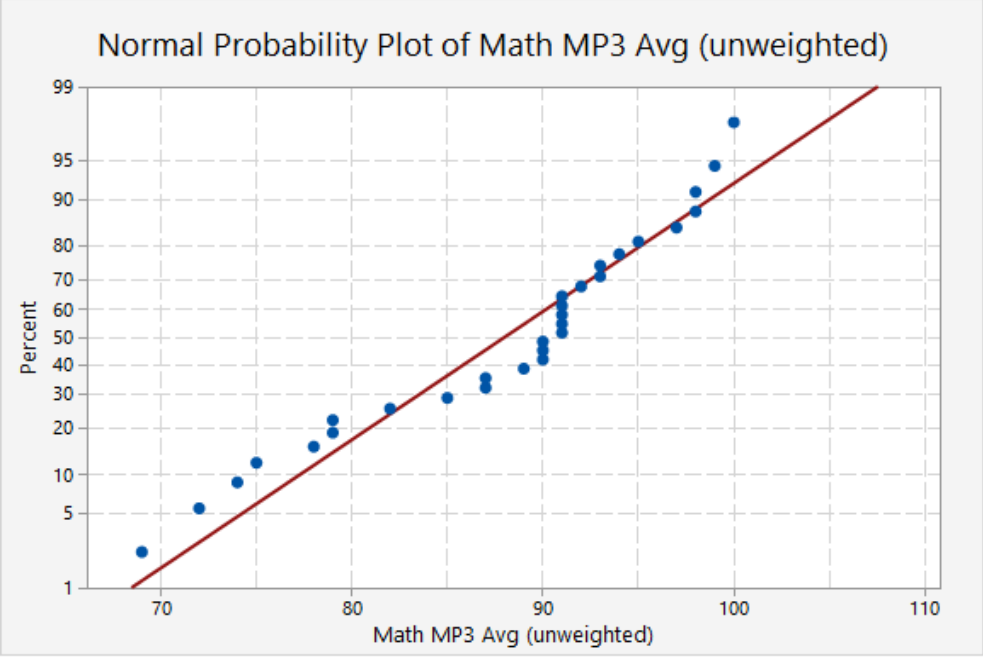


Figure 3.2: Distribution of MP3 Math Unweighted Averages, NPP



These graphs show that the distribution of the unweighted marking period 3 average in math, as reported by the survey respondents, have 0 outliers. This results in the median being 90.500 and the mean being 88. Clearly, the graphs show moderate left-skewness of the distribution and the mean being lower than the median should result in the median of 90.500 being the center of this distribution. There is greater spread in this distribution as Minitab Express calculated the standard deviation of these data to be 8.400. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

Science – What was your Marking Period 3 average in Science?

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
Science MP3 Avg (unweighted)	28	89.964	1.100	5.821	33.888	72.000	87.250	90.000	93.000	98.000	26.000	5.750	90

Figure 4.1: Distribution of MP3 Science Unweighted Averages, Boxplot

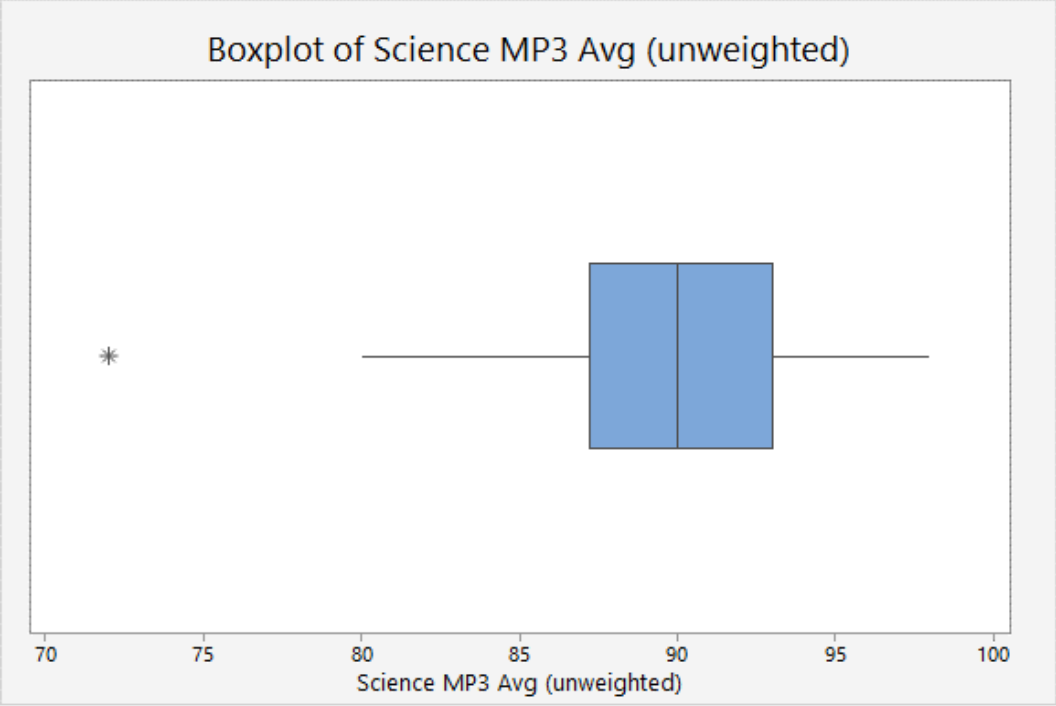
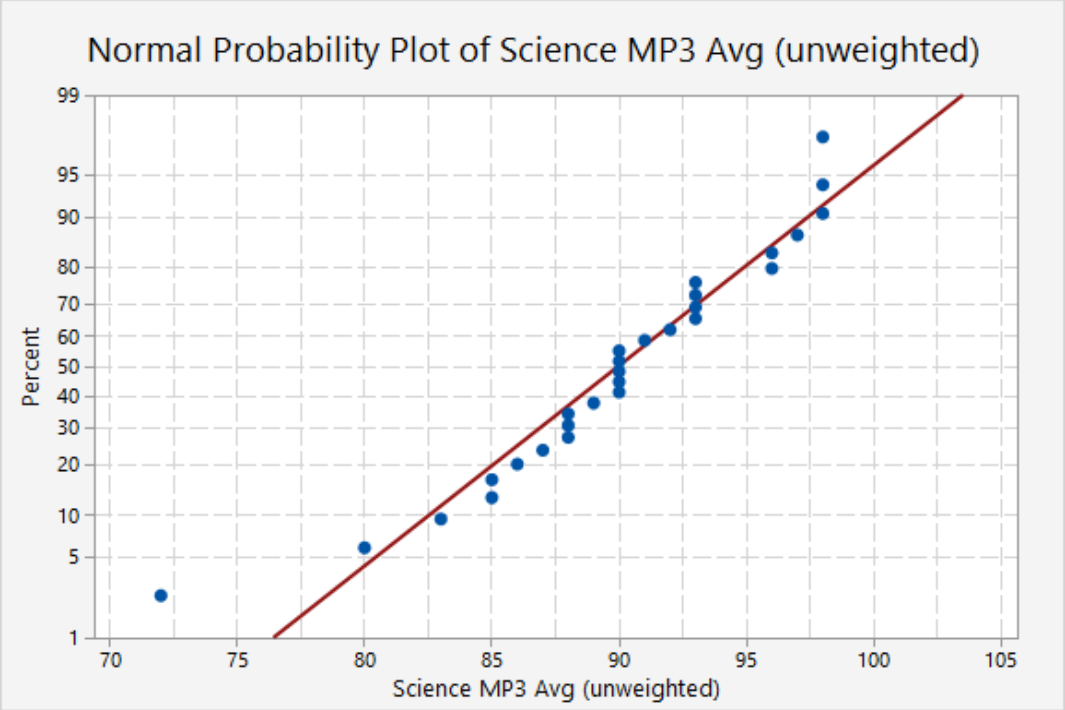


Figure 4.2: Distribution of MP3 Science Unweighted Averages, NPP



These graphs show that the distribution of the unweighted marking period 3 average in science, as reported by the survey respondents, have 1 outlier of 72. This results in the median being 90 and the mean being 89.964. Clearly, the graphs show slight left-skewness of the distribution and the mean being slightly lower than the median should result in the median of 90 being the center of this distribution. There is moderate spread in this distribution as Minitab Express calculated the standard deviation of these data to be 5.821. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

Average Unweighted Core Subject Performance – (calculated, not directly obtained)

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
Avg unweighted core performance	30	89.5722	0.8808	4.8243	23.2743	77.2500	86.5000	90.4583	93.4375	97.0000	19.7500	6.9375	86.5

Figure 1.1: Distribution of MP3 Average Unweighted Core Subject Performance, Boxplot

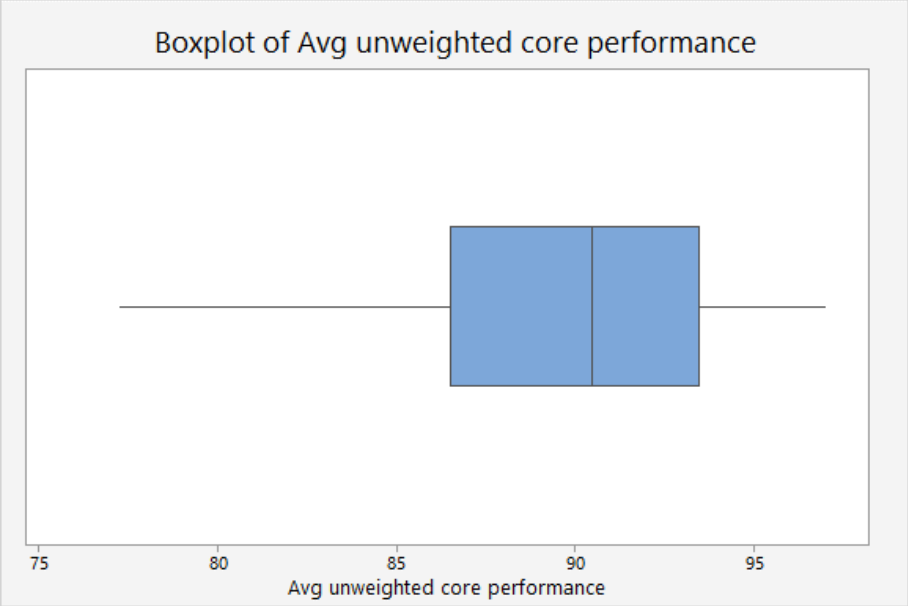
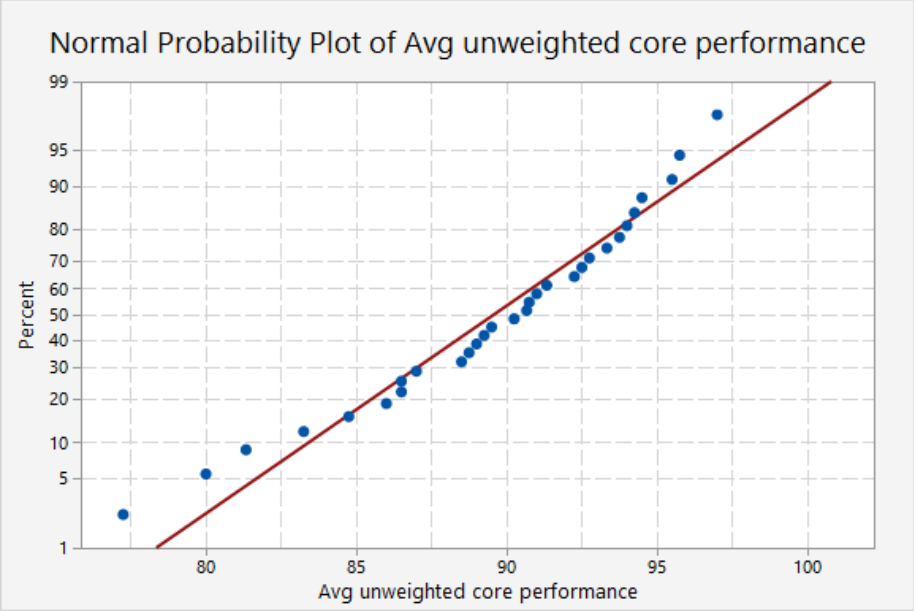


Figure 1.2: Distribution of MP3 Average Unweighted Core Subject Performance, NPP



These graphs show that the distribution of the average unweighted marking period 3 core subject performance, as calculated based on survey respondent-reported grades, have 0 outliers. This results in the median being 90.4583 and the mean being 89.5722. Clearly, the graphs show

moderate left-skewness of the distribution and the mean being moderately lower than the median should result in the median of 90.4853 being the center of this distribution. There is slight spread in this distribution as Minitab Express calculated the standard deviation of these data to be 4.8243. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

Average Weighted Core Subject Performance – (calculated, not directly obtained)

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
Avg weighted core performance	30	92.583	1.359	7.446	55.441	77.250	86.875	92.375	98.250	107.750	30.500	11.375	90.25, 96.00, 100.00

Figure 1.1: Distribution of MP3 Average Weighted Core Subject Performance, Boxplot

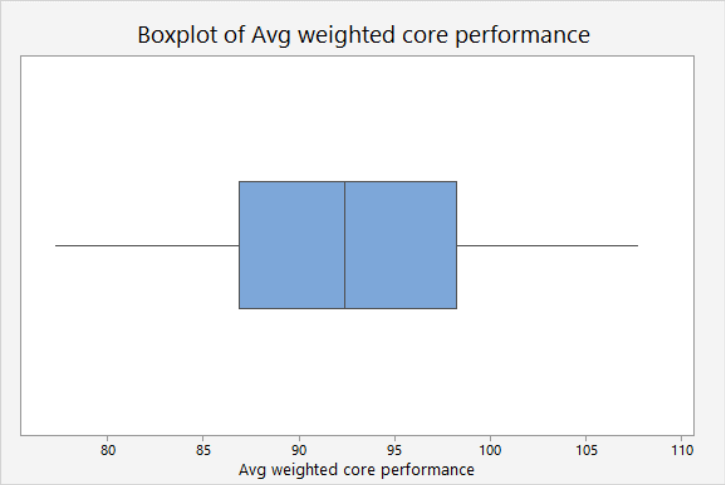
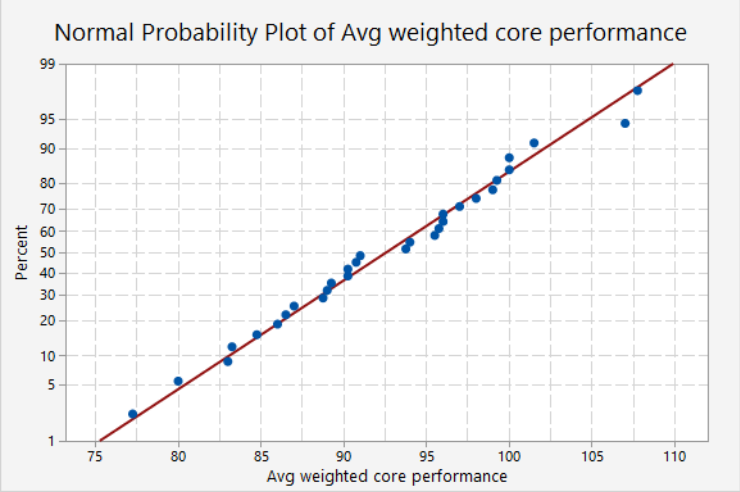


Figure 1.2: Distribution of MP3 Average Weighted Core Subject Performance, NPP



These graphs show that the distribution of the average weighted marking period 3 core subject performance, as calculated based on survey respondent-reported grades, have 0 outliers. This results in the median being 92.375 and the mean being 92.583. Clearly, the graphs show an approximately symmetrical distribution and the mean being only very slightly greater than the median can result in the mean of 92.583 being the center of this distribution. There is greater spread in this distribution as Minitab Express calculated the standard deviation of these data to be 7.446. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

Daily Music Listening – On average, about how many hours a day do you listen to music?

Variable	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
Daily Music Listening, hours	30	4.1167	0.6642	3.6381	13.2359	1.0000	2.0000	3.0000	5.0000	18.0000	17.0000	3.0000	3

Figure 1.1: Distribution of Daily Music Listening, Boxplot

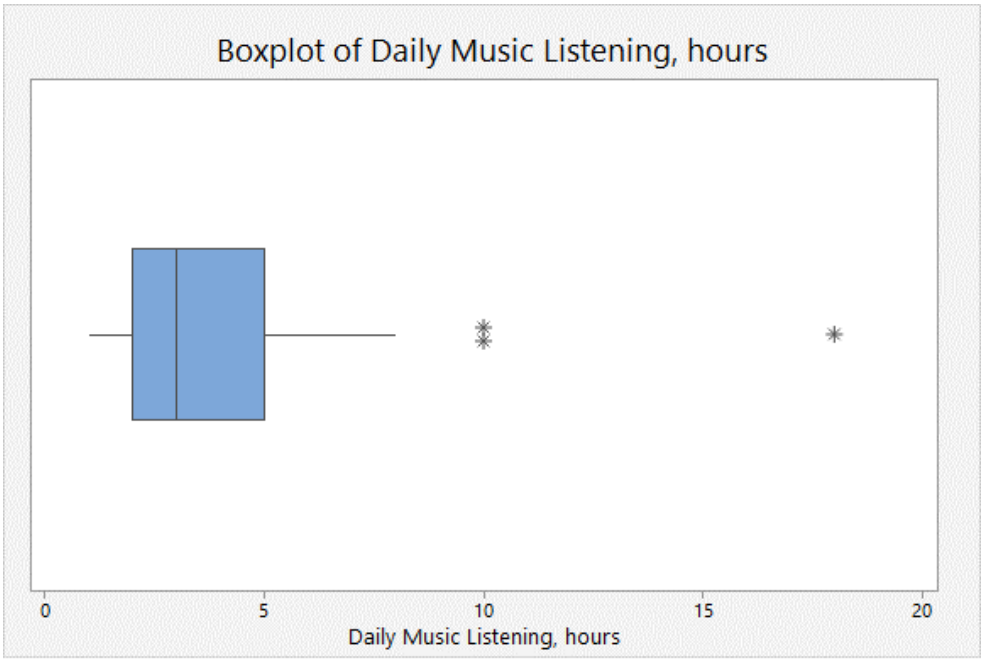
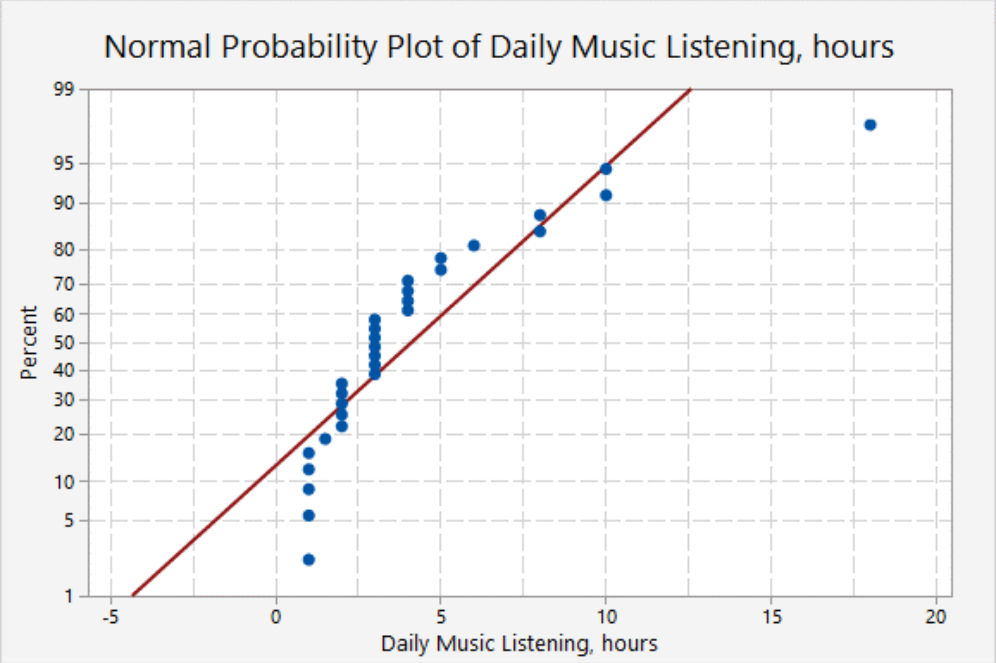


Figure 1.2: Distribution of Daily Music Listening, NPP



These graphs show that the distribution of daily music listening (in hours), as reported by the survey respondents, have 3 outliers—2 people listen to music for 10 hours a day and 1 person listens to music for 18 hours a day. This results in the median being 3 hours and the mean being 4.1167 hours. Clearly, the graphs show a greatly right-skewed distribution and the mean being only greater than the median should result in the median of 3 hours being the center of this distribution. There is greater spread in this distribution as Minitab Express calculated the standard deviation of these data to be 3.6381. The NPP shows a non-linear trend, indicating that the data are not normally distributed.

Comparing Numerical Distribution Using a Categorical Variable

Average Unweighted Core Subject Performance for Under- and Upper-classmen[2]

Side-by-Side Boxplots

Let "Under" represent the label "underclassman" and "Upper" represent the label "upperclassman".

Figure 1: Side-by-Side Boxplots of Average Unweighted Core Subject Performance for Under- and Upper-classmen

Variable	Under/Upper	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
Avg unweighted core performance	Under	16	89.7500	0.9134	3.6538	13.3500	83.2500	86.6250	89.3750	92.6875	95.5000	12.2500	6.0625	*
	Upper	14	89.369	1.613	6.036	36.432	77.250	85.208	90.875	93.813	97.000	19.750	8.604	*

The boxplots show that both distributions of average unweighted core subject performance for under- and upper-classmen contain no outliers. This results in the median average unweighted core subject performance for under-classmen being 89.3750 and the mean being 89.7500, and for the upper-classmen the median is 90.875 and the mean is 89.369. Clearly, the upper-classmen's

[2] Underclassmen are defined as students in grades 9 or 10 and upperclassmen are defined as students in grades 11 or 12.

mean of 89.369 is slightly lower than the under-classmen's mean of 89.7500 while the median for upper-classmen is greater than that of the under-classmen (90.875 > 89.3750). Because the distribution of data for upper-classmen is greatly left-skewed when compared to that of the under-classmen, the median unweighted averages of both groups should be used as the centers of the distributions. Thus, the higher center (median) for upper-classmen can be, in part, due to them possessing more skills and education (that they have learned for a greater number of years) than the under-classmen.

Clearly, there is significantly more variability in the distribution of data for upper-classmen than for under-classmen. This is shown by an almost doubled value of standard deviation for the upper-classmen when compared to the under-classmen (6.036 v. 3.6538). It is not quite clear what may be responsible for this greater variability.

Hypothesis Testing

A 2-sample t-test for the difference of sample means will be conducted to test if there is a difference between the sample means (of average unweighted core subject performance) of the under- and upper-classmen.

Let μ_1 represent the population mean unweighted core subject performance for under-classmen and μ_2 represent the population mean unweighted core subject performance for upper-classmen.

Method

μ_1: mean of Avg unweighted core performance when Under/Upper = Under
μ_2: mean of Avg unweighted core performance when Under/Upper = Upper
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 \neq 0$

$\alpha = 0.05$

Conditions:

1. **SRS**- each sample was randomly selected from the population of students taking at least 1 mathematics class this year using a random number generator
2. **Normality**- $n_1 = 16 < 30$ and $n_2 = 14 < 30$ which means that the sampling distributions cannot be considered approximately normal according to the Central Limit Theorem. However, the side-by-side boxplots show that there are no outliers in either distribution and the under-classmen distribution is barely skewed. Additionally, though the upper-classmen distribution is significantly skewed, one can still cautiously proceed with carrying out this test.
3. **Independence**- $10n_1 <? N_1$ $\quad 10n_2 <? N_2$

   $10 * 16 <? N_1$ $\quad 10 * 14 <? N_2$

   $160 < N_1$ $\quad 140 < N_2$

   It can be reasonably assumed that there are more than 160 under-classmen and 140 upper-classmen enrolled in a math class this year. Therefore, the independence condition is satisfied for each population and sample.

Formula:

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

T-Value	DF	P-Value
0.21	20	0.8393

Conclusion:

Because the p-value of the test = 0.8393 > α = 0.05, the null hypothesis cannot be rejected, meaning that there may be no significant difference between the true population mean unweighted core subject performance between under- and upper-classmen.

Average Weighted Core Subject Performance for Under- and Upper-classmen

Side-by-Side Boxplots

Let "Under" represent the label "underclassman" and "Upper" represent the label "upperclassman".

Figure 1: Side-by-Side Boxplots of Average Weighted Core Subject Performance for Under- and Upper-classmen

Variable	Under/Upper	N	Mean	SE Mean	StDev	Variance	Minimum	Q1	Median	Q3	Maximum	Range	IQR	Mode
Avg weighted core performance	Under	16	91.156	1.226	4.902	24.032	83.250	87.438	90.250	95.688	99.250	16.000	8.250	90.25
	Upper	14	94.214	2.544	9.517	90.576	77.250	85.625	96.500	100.375	107.750	30.500	14.750	100

The boxplots show that both distributions of average weighted core subject performance for under- and upper-classmen contain no outliers. This results in the median average weighted

core subject performance for under-classmen being 90.250 and the mean being 91.156, and for the upper-classmen the median is 96.500 and the mean is 94.214. Clearly, the upper-classmen's mean of 94.214 is greatly higher than the under-classmen's mean of 91.156 while the median for upper-classmen is even more greatly higher than that of the under-classmen (96.500 > 90.250). Because the distribution of data for upper-classmen is greatly left-skewed when compared to that of the under-classmen and the under-classmen data being moderately right-skewed when compared to that of the upper-classmen, the median unweighted averages of both groups should be used as the centers of the distributions. Thus, the higher center (median) for upper-classmen can be, in part, due to them possessing more skills and education (that they have learned for a greater number of years) than the under-classmen and also they have more opportunities to enroll in AP/IB classes which offer greater weight (+10 points) to their GPA.

Clearly, there is significantly more variability in the distribution of data for upper-classmen than for under-classmen. This is shown by an almost doubled value of standard deviation for the upper-classmen when compared to the under-classmen (9.517 v. 4.902). It is not quite clear what may be responsible for this greater variability though the upper-classmen taking a wider variety of differently weighted (CPB/CPA vs AP/IB) classes may be responsible for this variability.

Hypothesis Testing

A 2-sample t-test for the difference of sample means will be conducted to test if there is a difference between the sample means (of average weighted core subject performance) of the under- and upper-classmen.

Let μ_1 represent the population mean weighted core subject performance for under-classmen and μ_2 represent the population mean weighted core subject performance for upper-classmen.

Method

μ_1: mean of Avg weighted core performance when Under/Upper = Under
μ_2: mean of Avg weighted core performance when Under/Upper = Upper
Difference: $\mu_1 - \mu_2$

Equal variances are not assumed for this analysis.

Test

Null hypothesis	H_0: $\mu_1 - \mu_2 = 0$
Alternative hypothesis	H_1: $\mu_1 - \mu_2 \neq 0$

$\alpha = 0.05$

Conditions:

1. **SRS-** each sample was randomly selected from the population of students taking at least 1 mathematics class this year using a random number generator
2. **Normality-** $n_1 = 16 < 30$ and $n_2 = 14 < 30$ which means that the sampling distributions cannot be considered approximately normal according to the Central Limit Theorem. However, the side-by-side boxplots show that there are no outliers in either distribution and the under-classmen distribution is barely skewed. Additionally, though the upper-classmen distribution is significantly skewed, one can still cautiously proceed with carrying out this test.
3. **Independence-** $10n_1 <? N_1$ $10n_2 <? N_2$
   $10 * 16 <? N_1$ $10 * 14 <? N_2$
   $160 < N_1$ $140 < N_2$

   It can be reasonably assumed that there are more than 160 under-classmen and 140 upper-classmen enrolled in a math class this year. Therefore, the independence condition is satisfied for each population and sample.

Formula:

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

T-Value	DF	P-Value
-1.08	18	0.2931

Conclusion:

Because the p-value of the test = 0.2931 > α = 0.05, the null hypothesis cannot be rejected, meaning that there may be no significant difference between the true population mean weighted core subject performance between under- and upper-classmen.

Because both 2-sample t-tests for the difference of two population means, i.e. the average core subject performance of under- and upper-classmen, resulted in the discovery that the null hypotheses cannot be rejected, this enables one to better determine that there is not a significant difference in academic performance between under- and upper-classmen. Thus, if the students' music listening habits were to predict their academic performance in the core subject area, this would remain true for under- and upper-classmen because there is no significant difference between their academic performances.

Categorical Variables

Grade Level and Music Listening Habits

Maximum Music Listening Season

Two-Way Table

	Fall	Spring	Summer	Winter	*Missing*[3]	Total
Grade 9	0	1	3	1	0	5
Grade 10	2	1	5	2	1	10
Grade 11	0	2	2	3	1	7
Grade 12	2	0	2	2	0	6
Total	4	4	12	8	*2*	28

Hypothesis Testing

A Chi-square test of association will be conducted between grade level and maximum music listening season.

Hypotheses:

H_0: There is no association between grade level and maximum music listening season.

H_a: There is an association between grade level and maximum music listening season.

$\alpha = 0.05$

Conditions:

1. **SRS**- the sample was randomly selected from the population of students enrolled in at least 1 mathematics class this year using a random number generator
2. **All expected counts are at least 1**- all expected counts are not at least 1

[3] Missing generally means that the survey respondent indicated that they usually listen to music for the same amount of time in all season, so there is no one season in which they listen to music the most compared to other seasons. Missing will not be considered in hypothesis testing.

3. **No more than 20% of the expected counts are less than 5**- all (100% of) expected counts are less than 5
4. **Independence**- $10n_1 <? N_1$

   $10(28) <? N_1$

   $280 < N_1$

There are more than 280 students in the population that are enrolled in at least 1 mathematics class this year. Therefore, the independence condition is satisfied.

Though 2 out of 4 conditions are not satisfied, the Chi-square test of association will still be conducted with caution.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Chi-Square Test

	Chi-Square	DF
Pearson	7.42	9
Likelihood Ratio	9.31	9

4 cell(s) with expected counts less than 1. Chi-square approximation probably invalid.
16 cell(s) with expected counts less than 5.

Summary

Input	
Distribution	Chi-Square
Degrees of freedom	9
Noncentrality parameter	0
Input value	7.42

Cumulative Probability

x	P(X ≤ x)
7.42	0.406522

p-value $= 0.406522$

Conclusion:

Because the p-value = 0.406522 > α = 0.05, do not reject the null. There is not sufficient evidence to conclude that there is an association between grade level and maximum music listening season.

Minimum Music Listening Season

Two-Way Table

	Fall	Spring	Summer	Winter	*Missing*[4]	Total
Grade 9	2	0	0	2	1	4
Grade 10	4	1	3	2	1	10
Grade 11	3	0	1	3	1	7
Grade 12	1	0	3	2	0	6
Total	10	1	7	9	*3*	27

Hypothesis Testing

A Chi-square test of association will be conducted between grade level and minimum music listening season.

Hypotheses:

H_0: There is no association between grade level and minimum music listening season.

H_a: There is an association between grade level and minimum music listening season.

$\alpha = 0.05$

Conditions:

1. **SRS**- the sample was randomly selected from the population of students enrolled in at least 1 mathematics class this year using a random number generator
2. **All expected counts are at least 1**- all expected counts are not at least 1

[4] Missing generally means that the survey respondent indicated that they usually listen to music for the same amount of time in all season, so there is no one season in which they listen to music the most compared to other seasons. Missing will not be considered in hypothesis testing.

3. **No more than 20% of the expected counts are less than 5**- all (100% of) expected counts are less than 5
4. **Independence**- $10n_1 <? N_1$

   $10(27) <? N_1$

   $270 < N_1$

There are more than 270 students in the population that are enrolled in at least 1 mathematics class this year. Therefore, the independence condition is satisfied.

Though 2 out of 4 conditions are not satisfied, the Chi-square test of association will still be conducted with caution.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Chi-Square Test

	Chi-Square	DF
Pearson	6.51	9
Likelihood Ratio	7.79	9

4 cell(s) with expected counts less than 1.
Chi-square approximation probably invalid.
16 cell(s) with expected counts less than 5.

Summary

Input	
Distribution	Chi-Square
Degrees of freedom	9
Noncentrality parameter	0
Input value	6.51

Cumulative Probability

x	P(X ≤ x)
6.51	0.312013

p-value $= 0.312013$

Conclusion:

Because the p-value = 0.312013 > α = 0.05, do not reject the null. There is not sufficient evidence to conclude that there is an association between grade level and minimum music listening season.

Genre of English Music

Two-Way Table

	Country/Folk	Heavy metal	Hip hop	Jazz	N/A[5]	Other	Rap	Rock	Total
Grade 9	0	0	1	0	1	2	1	0	5
Grade 10	1	0	0	1	1	2	4	2	11
Grade 11	2	0	1	0	0	2	3	0	8
Grade 12	0	1	0	0	0	1	2	2	6
Total	3	1	2	1	2	7	10	4	30

Hypothesis Testing

A Chi-square test of association will be conducted between grade level and genre of English music that is most listened to by the participants on average.

Hypotheses:

H_0: There is no association between grade level and genre of English music.

H_a: There is an association between grade level and genre of English music.

$\alpha = 0.05$

[5] N/A means that the participant responded that they listen to more non-English music than English music on the average, so the genre of English music is not applicable.

Conditions:

1. **SRS**- the sample was randomly selected from the population of students enrolled in at least 1 mathematics class this year using a random number generator
2. **All expected counts are at least 1**- 21 expected counts are not at least 1
3. **No more than 20% of the expected counts are less than 5**- all (100% of) expected counts are less than 5
4. **Independence**- $10n_1 <? N_1$

   $10(30) <? N_1$

   $300 < N_1$

There are more than 300 students in the population that are enrolled in at least 1 mathematics class this year. Therefore, the independence condition is satisfied.

Though 2 out of 4 conditions are not satisfied, the Chi-square test of association will still be conducted with caution.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Chi-Square Test

	Chi-Square	DF
Pearson	18.79	21
Likelihood Ratio	21.02	21

21 cell(s) with expected counts less than 1. Chi-square approximation probably invalid.
32 cell(s) with expected counts less than 5.

Summary

Input	
Distribution	Chi-Square
Degrees of freedom	21
Noncentrality parameter	0
Input value	18.79

Cumulative Probability

x	P(X ≤ x)
18.79	0.401391

p-value = 0.401391

Conclusion:

Because the p-value = 0.401391 > α = 0.05, do not reject the null. There is not sufficient evidence to conclude that there is an association between grade level and genre of English music that is most listened to by participants.

Music Listening Habits and Average Weighted Core Subject Performance

Maximum Music Listening Season

Two-Way Table

	100-104	105-110	75-79	80-84	85-89	90-94	95-99	Total
Fall	0	0	0	1	2	0	1	4
Spring	0	0	1	0	0	1	2	4
Summer	1	1	0	2	2	3	3	12
Winter	2	0	0	1	2	1	2	8
Missing[6]	0	1	0	0	0	1	0	*2*
Total	3	1	1	4	6	5	8	28

Hypothesis Testing

A Chi-square test of association will be conducted between maximum music listening season and average weighted core subject performance.

Hypotheses:

H_0: There is no association between maximum music listening season and average weighted core subject performance.

[6] Missing generally means that the survey respondent indicated that they usually listen to music for the same amount of time in all season, so there is no one season in which they listen to music the most compared to other seasons. Missing will not be considered in hypothesis testing.

H_a: There is an association between maximum music listening season and average weighted core subject performance.

$\alpha = 0.05$

Conditions:

1. **SRS**- the sample was randomly selected from the population of students enrolled in at least 1 mathematics class this year using a random number generator
2. **All expected counts are at least 1**- 17 expected counts are not at least 1
3. **No more than 20% of the expected counts are less than 5**- all (100% of) expected counts are less than 5
4. **Independence**- $10n_1 <? N_1$

   $10(28) <? N_1$

   $280 < N_1$

There are more than 280 students in the population that are enrolled in at least 1 mathematics class this year. Therefore, the independence condition is satisfied.

Though 2 out of 4 conditions are not satisfied, the Chi-square test of association will still be conducted with caution.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Chi-Square Test

	Chi-Square	DF
Pearson	15.34	18
Likelihood Ratio	15.56	18

17 cell(s) with expected counts less than 1. Chi-square approximation probably invalid.
28 cell(s) with expected counts less than 5.

Summary

Input

Distribution	Chi-Square
Degrees of freedom	18
Noncentrality parameter	0
Input value	15.34

Cumulative Probability

x	P(X ≤ x)
15.34	0.361496

p-value = 0.361496

Conclusion:

Because the p-value = 0.361496 > α = 0.05, do not reject the null. There is not sufficient evidence to conclude that there is an association between maximum music listening season and average weighted core subject performance.

Minimum Music Listening Season

Two-Way Table

	100-104	105-110	75-79	80-84	85-89	90-94	95-99	Total
Fall	2	1	1	0	2	1	3	10
Spring	0	0	0	0	1	0	0	1
Summer	0	0	0	1	3	0	3	7
Winter	1	0	0	3	0	3	2	9
Missing[7]	0	1	0	0	0	2	0	*3*
Total	3	1	1	4	6	4	8	27

Hypothesis Testing

A Chi-square test of association will be conducted between minimum music listening season and average weighted core subject performance.

[7] Missing generally means that the survey respondent indicated that they usually listen to music for the same amount of time in all season, so there is no one season in which they listen to music the most compared to other seasons. Missing will not be considered in hypothesis testing.

Hypotheses:

H_0: There is no association between minimum music listening season and average weighted core subject performance.

H_a: There is an association between minimum music listening season and average weighted core subject performance.

$\alpha = 0.05$

Conditions:

1. **SRS-** the sample was randomly selected from the population of students enrolled in at least 1 mathematics class this year using a random number generator
2. **All expected counts are at least 1-** 14 expected counts are not at least 1
3. **No more than 20% of the expected counts are less than 5-** all (100% of) expected counts are less than 5
4. **Independence-** $10n_1 <? N_1$

   $10(27) <? N_1$

   $270 < N_1$

There are more than 270 students in the population that are enrolled in at least 1 mathematics class this year. Therefore, the independence condition is satisfied.

Though 2 out of 4 conditions are not satisfied, the Chi-square test of association will still be conducted with caution.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Chi-Square Test

	Chi-Square	DF
Pearson	19.10	18
Likelihood Ratio	22.86	18

14 cell(s) with expected counts less than 1.
Chi-square approximation probably invalid.
28 cell(s) with expected counts less than 5.

Summary

Input	
Distribution	Chi-Square
Degrees of freedom	18
Noncentrality parameter	0
Input value	19.1

Cumulative Probability

x	P(X ≤ x)
19.1	0.614310

p-value = 0.614310

Conclusion:

Because the p-value = 0.614310 > α = 0.05, do not reject the null. There is not sufficient evidence to conclude that there is an association between minimum music listening season and average weighted core subject performance.

Music Language

Two-Way Table

	100-104	105-110	75-79	80-84	85-89	90-94	95-99	Total
English	3	2	1	4	5	5	8	28
Non-English	0	0	0	0	1	1	0	2
Total	3	2	1	4	6	6	8	30

Hypothesis Testing

A Chi-square test of association will be conducted between music language and average weighted core subject performance.

Hypotheses:

H_0: There is no association between music language and average weighted core subject performance.

H_a: There is an association between music language and average weighted core subject performance.

$\alpha = 0.05$

Conditions:

1. **SRS-** the sample was randomly selected from the population of students enrolled in at least 1 mathematics class this year using a random number generator
2. **All expected counts are at least 1-** 8 expected counts are not at least 1
3. **No more than 20% of the expected counts are less than 5-** about 21% expected counts are less than 5
4. **Independence-** $10n_1 <? N_1$

   $10(30) <? N_1$

   $300 < N_1$

There are more than 300 students in the population that are enrolled in at least 1 mathematics class this year. Therefore, the independence condition is satisfied.

Though 2 out of 4 conditions are not satisfied, the Chi-square test of association will still be conducted with caution.

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Chi-Square Test

	Chi-Square	DF
Pearson	3.21	6
Likelihood Ratio	3.88	6

8 cell(s) with expected counts less than 1.
Chi-square approximation probably invalid.
11 cell(s) with expected counts less than 5.

Summary

Input	
Distribution	Chi-Square
Degrees of freedom	6
Noncentrality parameter	0
Input value	3.21

Cumulative Probability

x	P(X ≤ x)
3.21	0.217934

p-value = 0.217934

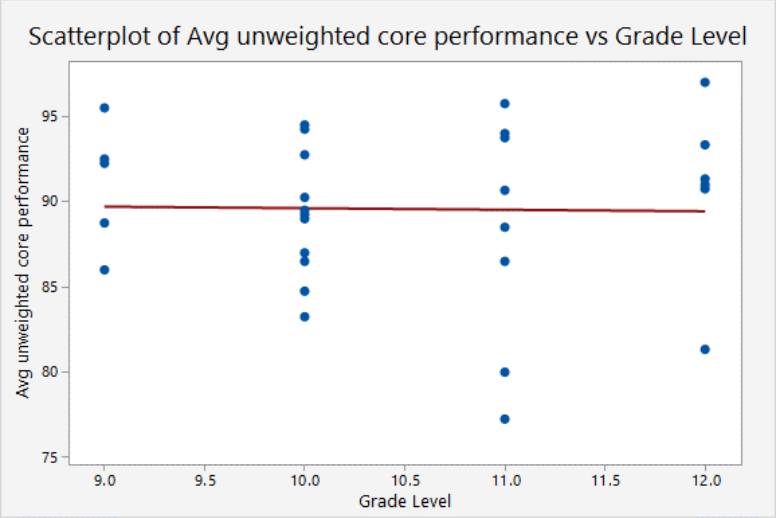
Conclusion:

Because the p-value = 0.214934 > α = 0.05, do not reject the null. There is not sufficient evidence to conclude that there is an association between music language and average weighted core subject performance.

Study of Bivariate Data

Grade Level and Average Core Academic Performance

Unweighted



Correlation

Pearson correlation of Grade Level and Avg unweighted core performance = -0.018898
P-Value = 0.9210

Model Summary

S	R-sq	R-sq(adj)
4.90886	0.04%	0.00%

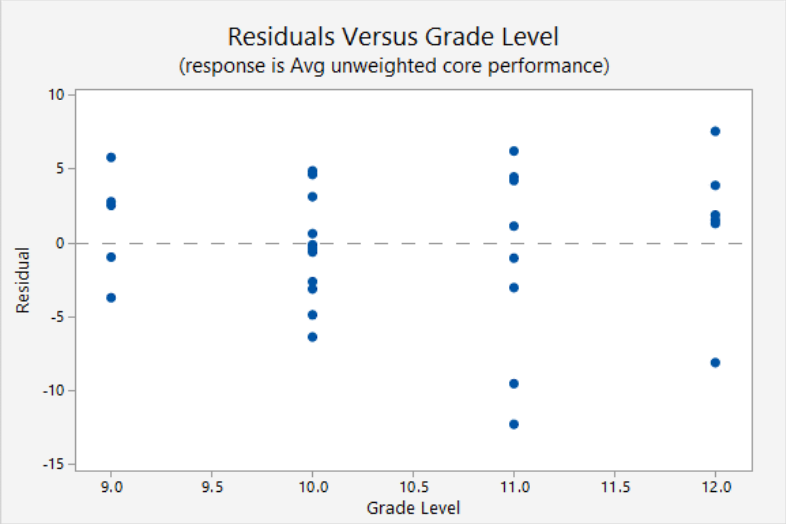
Regression Equation

Avg unweighted core performance = 90.521 − 0.0904 Grade Level

Fits and Diagnostics for Unusual Observations

Obs	Avg unweighted core performance	Fit	Resid	Std Resid	
5	77.25	89.5270	-12.2770	-2.55	R

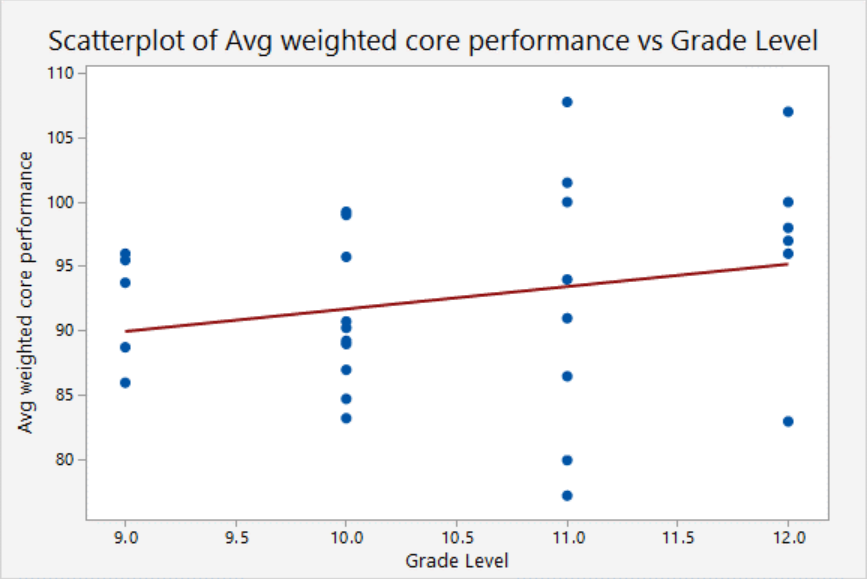
R Large residual



Discussion

The explanatory variable was grade level, a discrete quantitative variable with positive integer values [9, 12]. The response variable was average unweighted core subject performance, a continuous quantitative variable with positive rational values. I expected these variables to have a slight positive linear relationship because though one would expect that as student's skills increase over the years in high school their academic performance would also increase, the level of difficulty of the classes would also rise proportionately, keeping a relative balance of overall academic performance throughout the years. Based on the scatterplot, there appears to be a very weak linear relationship, overall, between grade level and average unweighted core subject performance. The r-value of -0.018898 shows that there is a very weak negative linear relationship between grade level and average unweighted core subject performance. This matches with the scatterplot's least-squares regression line. There appears to be more of a random scattering of residuals than a set pattern upon closer examination of the residual plot, so a linear model may be appropriate. There appears to be 1 outlier, denoted by Minitab as R Large residual, of y-value = 77.25. This shows that that student's average unweighted core subject performance is very unlike the other students' averages in the sample.

Weighted



Correlation

Pearson correlation of Grade Level and Avg weighted core performance = 0.235325
P-Value = 0.2106

Model Summary

S	R-sq	R-sq(adj)
7.36487	5.54%	2.16%

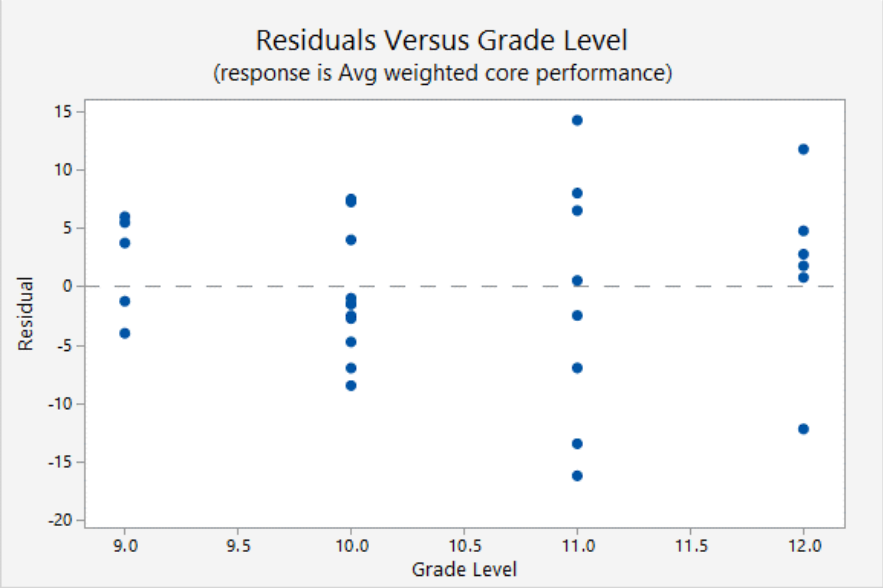
Regression Equation

Avg weighted core performance = 74.34 + 1.737 Grade Level

Fits and Diagnostics for Unusual Observations

Obs	Avg weighted core performance	Fit	Resid	Std Resid	
5	77.25	93.4520	-16.2020	-2.25	R

R Large residual

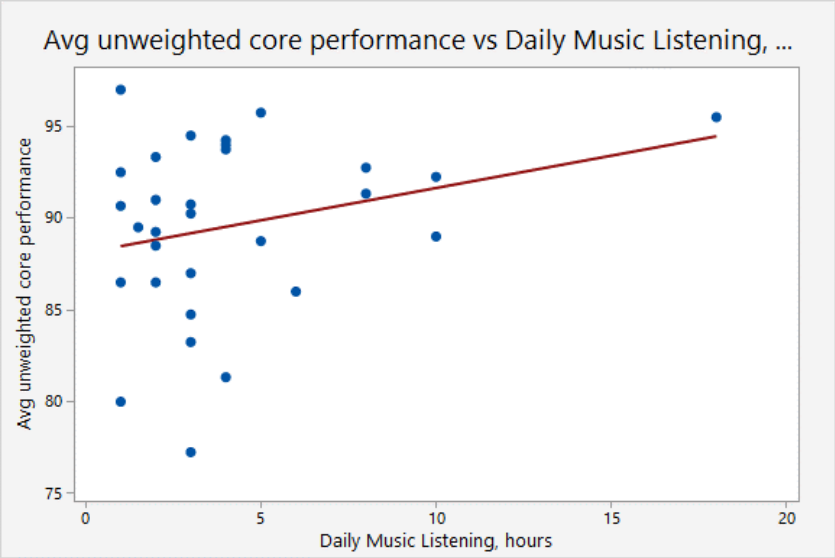


Discussion

The explanatory variable was grade level, a discrete quantitative variable with positive integer values [9, 12]. The response variable was average weighted core subject performance, a continuous quantitative variable with positive rational values. I expected these variables to have a moderate positive linear relationship because students in higher grade levels oftentimes take more heavily weighted classes like Honors, AP and IB. This would add more points to their averages, thereby increasing it as the years increase. Based on the scatterplot, there appears to be a slight linear relationship, overall, between grade level and average weighted core subject performance. The r-value of 0.235325 shows that there is a weak positive linear relationship between grade level and average weighted core subject performance. This matches with the scatterplot's least-squares regression line. There appears to be a slight "fan-shape" upon closer examination of the residual plot, so a linear model may not be appropriate. There appears to be 1 outlier, denoted by Minitab as R Large residual, of y-value = 77.25. This shows that that student's average weighted core subject performance is very unlike the other students' averages in the sample.

Daily Music Listening and Average Core Academic Performance

Unweighted



Correlation

Pearson correlation of Daily Music Listening, hours and Avg unweighted core performance = 0.265551
P-Value = 0.1561

Model Summary

S	R-sq	R-sq(adj)
4.73346	7.05%	3.73%

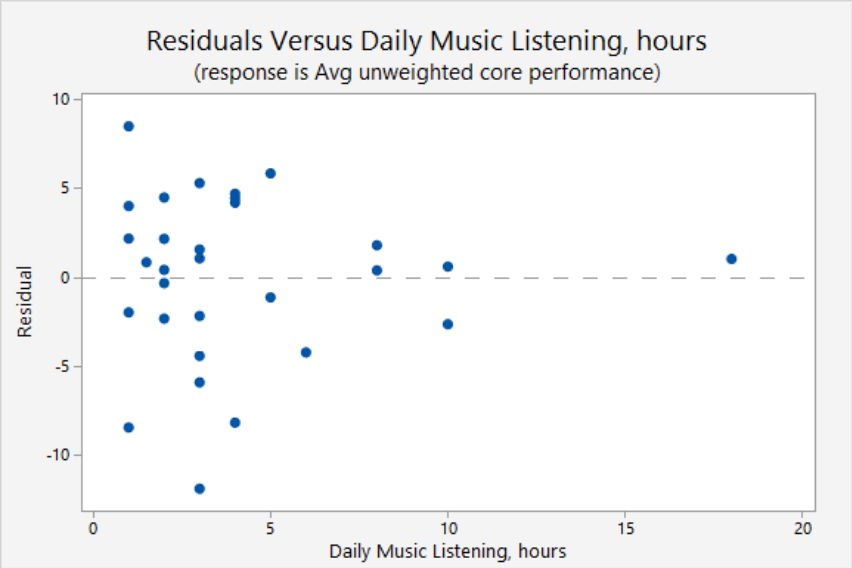
Fits and Diagnostics for Unusual Observations

Obs	Avg unweighted core performance	Fit	Resid	Std Resid	
5	77.25	89.1790	-11.9290	-2.57	R
11	95.5	94.4610	1.0390	0.32	X

R Large residual
X Unusual X

Regression Equation

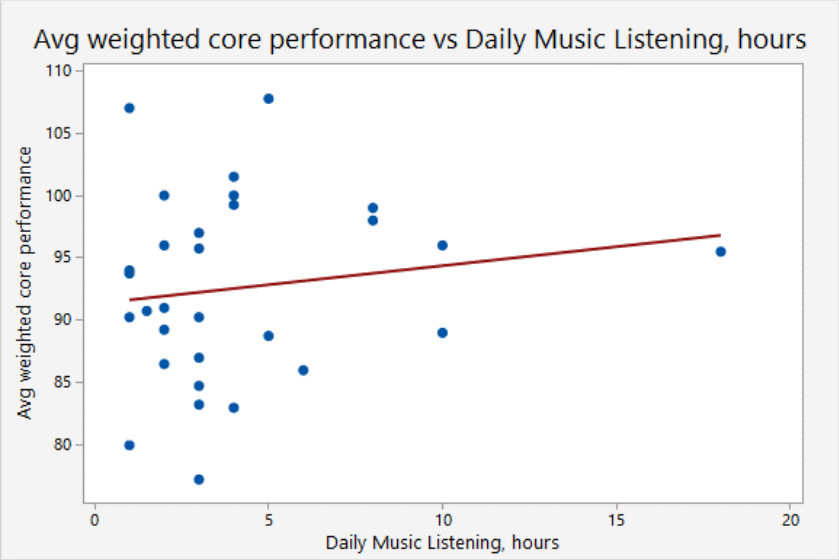
Avg unweighted core performance = 88.123 + 0.3521 Daily Music Listening, hours



Discussion

The explanatory variable was daily music listening (in hours), a continuous quantitative variable with positive integer values. The response variable was average unweighted core subject performance, a continuous quantitative variable with positive rational values. I expected these variables to have a moderate positive linear relationship because other psychological studies showed that listening to certain types of music further encourages cortical development, and therefore, can result in stronger academic performance. Based on the scatterplot, there appears to be a slight linear relationship, overall, between daily music listening hours and average unweighted core subject performance. The r-value of 0.265551 shows that there is a weak positive linear relationship between daily music listening hours and average unweighted core subject performance. This matches with the scatterplot's least-squares regression line. There appears to be a strong "fan-shape" upon closer examination of the residual plot, so a linear model may not be appropriate. There appears to be 1 outlier, denoted by Minitab as R Large residual, of y-value = 77.25 and 1 influential point with values (18, 95.5). This shows that that student's average weighted core subject performance is very unlike the other students' averages in the sample and that the other student's daily music listening hours is very unlike the other students' hours in the sample.

Weighted



Correlation

Pearson correlation of Daily Music Listening, hours and Avg weighted core performance = 0.148722
P-Value = 0.4328

Model Summary

S	R-sq	R-sq(adj)
7.49340	2.21%	0.00%

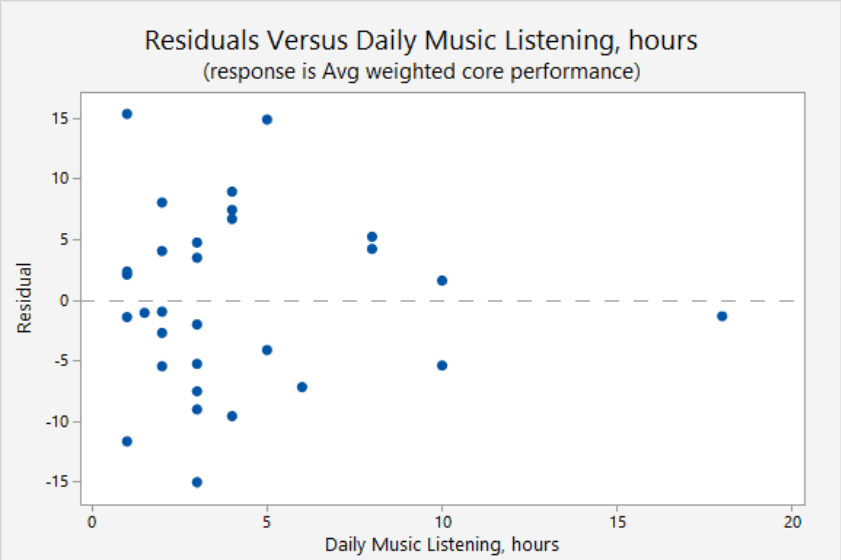
Fits and Diagnostics for Unusual Observations

Obs	Avg weighted core performance	Fit	Resid	Std Resid		
5	77.25	92.2434	-14.9934	-2.04	R	
11	95.5	96.8091	-1.3091	-0.26		X
29	107.75	92.8522	14.8978	2.02	R	
30	107	91.6347	15.3653	2.11	R	

R Large residual
X Unusual X

Regression Equation

Avg weighted core performance = 91.330 + 0.3044 Daily Music Listening, hours



Discussion

The explanatory variable was daily music listening (in hours), a continuous quantitative variable with positive integer values. The response variable was average weighted core subject performance, a continuous quantitative variable with positive rational values. I expected these variables to have a moderate positive linear relationship because other psychological studies showed that listening to certain types of music further encourages cortical development, and therefore, can result in stronger academic performance. Based on the scatterplot, there appears to be a slight linear relationship, overall, between daily music listening hours and average unweighted core subject performance. The r-value of 0.148722 shows that there is a weak positive linear relationship between daily music listening hours and average weighted core subject performance. This matches with the scatterplot's least-squares regression line. There appears to be a strong "fan-shape" upon closer examination of the residual plot, so a linear model may not be appropriate. There appears to be 3 outliers, denoted by Minitab as R Large residual, of y-values = 77.25, 107.75, 107 and 1 influential point with values (18, 95.5). This shows that those student's average weighted core subject performance is very unlike the other students' averages in the sample and that the other student's daily music listening hours is very unlike the other students' hours in the sample.

Conclusion

This research project concludes that a person's music listening habits do not predict their academic performance in a core subject area.

The analysis of all of the numerical variables resulted in non-normal data distributions. In the case of the number of classes of a particular level, it should be expected that more people would take CPA classes because that level is considered to be the average or the academic norm while CPB, Honors and AP/IB classes are considered more extreme. However, this was not the case. Nonetheless, the unweighted and weighted grades in each of the core classes were somewhat performance were not normally distributed but did not have any strong skewness. The Analysis of Numerical Variables shows that the data are not evenly spread in a normal distribution, thereby making it more difficult to draw conclusions about them.

The two-sample t-tests showed that there were no significant differences between the core academic subject performance (weighted and unweighted) between under- and upper-classmen. This showed that under- and upper-classmen's academic performances in core subjects can be looked at together and do not need to be looked at individually.

The Chi-square tests of association between various elements of music listening habits and core academic subject performances resulted in the failure to reject the null, meaning that there was not sufficient evidence to conclude that there was an association between those elements of music listening habits and core academic subject performances.

These failings could be due to a couple of flaws in the survey process. For example, a couple of the survey respondents responded that they mostly listened to non-English music, and therefore, did not answer the music genre question. This resulted in $n < 30$ for the answers to that question, failing to meet the Central Limit Theorem requirements. This could have been resolved by having a higher initial sample. Another flaw in the survey process could have been the fact that

some respondents circled more than one genre for the genre of English music they listened to the most. This could have been fixed by more strongly emphasizing the need to circle ***one*** genre.

In conclusion, this research project finds that a person's music listening habits do not predict their academic performance in a core subject area.

Appendices

Appendix 1 — Raw Data

Grade Level	CPB Classes	CPA Classes	Honors Classes	AP/IB Classes	History Class Level	History MP3 Avg (unweighted)	History MP3 Avg (weighted)	English Class Level	English MP3 Avg (unweighted)	English MP3 Avg (weighted)	Math Class Level	Math MP3 Avg (unweighted)	Math MP3 Avg (weighted)	Science Class Level	Science MP3 Avg (unweighted)	Science MP3 Avg (weighted)	Daily Music Listening, hours	Spring	Summer	Fall	Winter	Max Music Listening Season	Min Music Listening Season	Music Language	Genre of English Music
10	3	0	1	0	CPB	90	90	Honors	85	90	CPB	90	90	CPB	93	93	1.5	2	3	1	4	Winter	Fall	NE	N/A
10	0	1	3	0	CPA	89	89	Honors	93	98	Honors	74	79	Honors	90	95	1	3	4	2	1	Summer	Winter	E	Country/Folk
10	4	0	0	0	CPB	89	89	CPB	78	78	CPB	98	98	CPB	92	92	2	3	1	4	2	Fall	Summer	E	Rock
10	0	0	4	1	Honors	93	98	Honors	96	101	Honors	91	96	AP/IB	91	101	8	2	4	1	3	Summer	Fall	E	Rap
11	2	2	0	0	CPA	79	79	CPA	86	86	CPB	72	72	CPB	72	72	3	4	3	1	2	Spring	Fall	E	Rap
12	0	4	0	3	AP/IB	93	103	AP/IB	95	105	CPA	78	78	CPA	98	98	2	2	1	4	3	Fall	Summer	E	Other
11	0	0	0	7	AP/IB	93	103	AP/IB	97	107	AP/IB	97	97	AP/IB	89	99	4	2	3	1	4	Winter	Fall	E	Other
11	1	6	0	0	CPA	80	80	CPA	70	70	CPB	90	90	CPA	80	80	1	3	4	2	1	Summer	Winter	E	Hip hop
11	0	6	0	0	CPA	89	89	CPA	93	93	CPA	79	79	CPA	85	85	2	2	1	3	4	Winter	Summer	E	Country/Folk
9	0	5	1	0	CPA	95	95	CPA	93	93	CPA	89	89	Honors	93	98	1	3	4	3	3	Summer		E	Rap
9	4	0	0	0	CPB	97	97	CPB	88	88	CPB	99	99	CPB	98	98	18	3	4	2	1	Summer	Winter	E	Hip hop
10	0	5	0	0	CPA	86	86	CPA	82	82	CPA	75	75	CPA	90	90	3	3	4	2	1	Summer	Winter	E	Rap
10	0	3	1	0	CPA	96	96	Honors	95	100	CPA	91	91	CPA	96	96	3	2	4	1	3	Summer	Fall	E	Other
12	0	2	2	2	AP/IB	92	102	CPA	95	95	AP/IB	93	103				2	2	4	3	1	Summer	Winter	E	Rock
11	0	4	0	1	CPA	88	88	CPA	91	91	AP/IB	82	92	CPA	93	93	2	3	4	2	1	Summer	Winter	E	Rap
10	0	4	0	0	CPA	90	90	CPA	85	85	CPA	90	90	CPA	83	83	3	2	1	4	3	Fall	Summer	E	Rap
11	0	2	1	2	AP/IB	86	96	CPA	94	94	CPA	92	92				1	4	2	3	1	Spring	Winter	E	Rap
12	0	3	2	2	AP/IB	90	100	CPA	90	90	Honors	95	100	AP/IB	88	98	3	2	1	3	4	Winter	Summer	E	Heavy metal
9	0	1	4	0	Honors	89	94	Honors	92	97	CPA	98	98	Honors	90	95	10	4	2	3	1	Spring	Winter	E	Other
10	0	0	5	2	Honors	96	101	Honors	91	96	Honors	93	98	Honors	97	102	4	4	3	1	2	Spring	Fall	E	Jazz
9	4	0	0	0	CPB	90	90	CPB	90	90	CPB	85	85	CPB	90	90	5	2	4	1	3	Summer	Fall	NE	N/A
11	0	1	3	1	AP/IB	96	106	AP/IB	96	106	Honors	87	92	CPA	96	96	4	2	3	1	4	Winter	Fall	E	Country/Folk
9	0	4	0	0	CPA	85	85	CPA	80	80	CPA	91	91	CPA	88	88	6	2	3	1	4	Winter	Fall	E	Other
10	0	6	0	0	CPA	85	85	CPA	90	90	CPA	79	79	CPA	85	85	3	2	1	3	4	Winter	Summer	E	Rap
10	4	0	0	0	CPB	87	87	CPB	85	85	CPB	91	91	CPB	93	93	10	1	4	2	3	Summer	Spring	E	Other
12	0	1	0	5				CPA	100	100	AP/IB	87	97	AP/IB	87	97	8	2	1	3	4	Winter	Summer	E	Rap
10	0	5	0	0	CPA	91	91	CPA	89	89	CPA	91	91	CPA	90	90	3	2	2	2	2			E	Rock
12	0	3	2	2				CPA	89	89	Honors	69	74	CPA	86	86	4	3	2	4	1	Fall	Winter	E	Rap
11	0	1	0	6	AP/IB	99	109	AP/IB	102	110	AP/IB	94	104	AP/IB	88	108	5	2	2	2	2			E	Other
12	0	0	0	8	AP/IB	93	103	AP/IB	97	107	AP/IB	100	110	AP/IB	98	108	1	2	4	1	3	Summer	Fall	E	Rock