

Statistical Analysis of Pitch Distributions Among Orchestral Instruments in Dvořák's  
Symphony No. 9, Movement IV.

IB Math

HL Internal Assessment (IA)

Alice Y. Hu

Word Count: 2569

## **1. Introduction**

Music and mathematics are similar in nature, both heavily relying on patterns and relationships. In orchestral compositions, composers use various pitch distributions for different instruments to create a balanced piece. Antonín Leopold Dvořák's Symphony No. 9 in E minor, Movement IV ("From the New World Symphony"), is exemplary of the use of patterns and blends among instruments to create a masterpiece. As a flutist and a mathematics student, I am interested in the mathematical explanation of how an accomplished composer uses pitch distributions across different instruments in an orchestra, the randomness of pitch selections when composing, and how a composer's choice of pitch distribution is heavily dependent on the range of a given instrument.

### **1.1 Aims of the Exploration**

This investigation will evaluate pitch distributions using statistical analysis across orchestral instruments in Dvořák's Symphony No. 9, Movement IV, for the instruments flute, violin, cello, B-flat clarinet, trumpet, trombone, timpani, and tuba. In order to fulfill this overall objective, a basic summary of central pitch tendency and dispersion (mean, median, range, standard deviation) will be collected to summarize the pitch of each instrument's data. To define pitch values quantitatively, they will be expressed as MIDI values. MIDI stands for Musical Instrument Digital Interface. It is "an American-Japanese technical standard that describes a communication protocol, digital interface, and electrical connectors that connect a wide variety of electronic musical instruments, computers, and related audio devices for playing, editing, and recording music" (Smith and Howell). The symmetry and shape of each distribution will be charted in a histogram

and compared using quantitative measures of skewness and kurtosis. Using the collected data, correlations will be calculated to determine a mathematical basis for differences, such as whether instruments with larger pitch ranges exhibit differences in skewness and/or kurtosis across their respective ranges. Finally, the ANOVA (Analysis of Variance) test will be used to determine whether the mean MIDI values across instrument groups (brass: trumpet, trombone, and tuba; woodwind: flute and B-flat clarinet; string: violin and cello) differ significantly.

## **1.2 Background Information**

Dvořák's Symphony No. 9 is one of the most frequently performed orchestral works in the world, praised for its effective use of melody and harmony and its rich instrumentation. Each instrument occupies a particular role in the orchestra: violins, trumpets, and flutes take melodic lines, trombone and cello offer counter melodies, and timpani and tuba provide bass lines. These instruments were selected for their contrasting roles in the symphony and for their different ranges. It ultimately allows for a meaningful comparison between the instrument range and its statistical distribution. This variety enables a meaningful comparison of how pitch usage and distribution differ among distinct instrument families: strings, woodwinds, and brass.

### **1.2.1 Pitch values**

As mentioned above, pitch quantification will be performed by converting pitches to MIDI values, which are continuous numerical representations of pitch. As the note rises in pitch, the MIDI value is higher. For simplicity, MIDI values will all be discrete numerical values derived from sheet music notes converted to MIDI.

Special ornamentations used in music, such as grace notes, cue notes, and trill notes, will be removed from the dataset. Gracenotes will be omitted due to their short length and impracticality for testing. Cue notes, which only indicate when another instrument or melodic line is performed, are not normally performed by the instrument, so they will be omitted in the study. Finally, trill notes involve the reiteration of a single note over a period of time. Because the number of reiterations varies across players, to ensure clarity, only the main note indicated by the trill will be included in the collected data. This allows the pitch distributions in the symphony to be described quantitatively and accurately, providing a possible way to determine the shape and spread.

### **1.2.2 Instrumentation Analyzed**

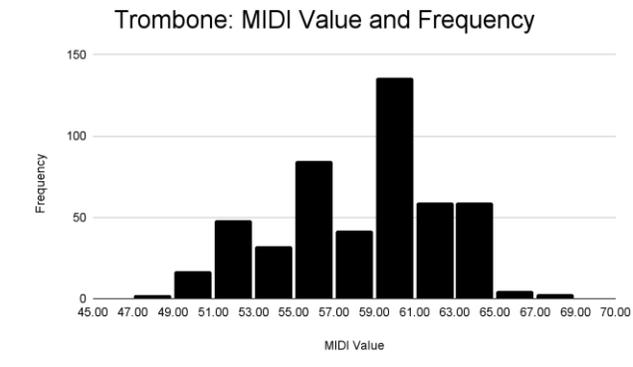
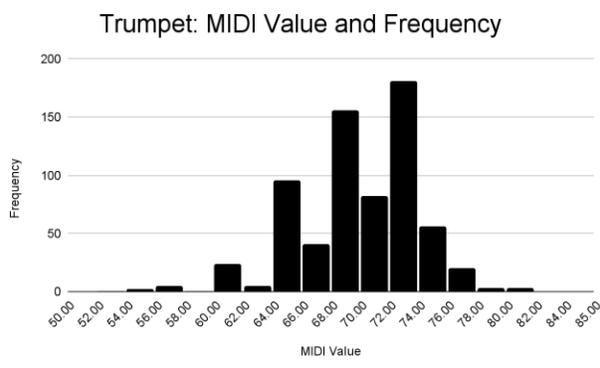
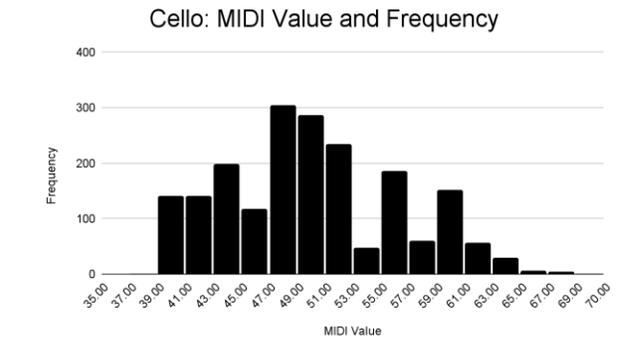
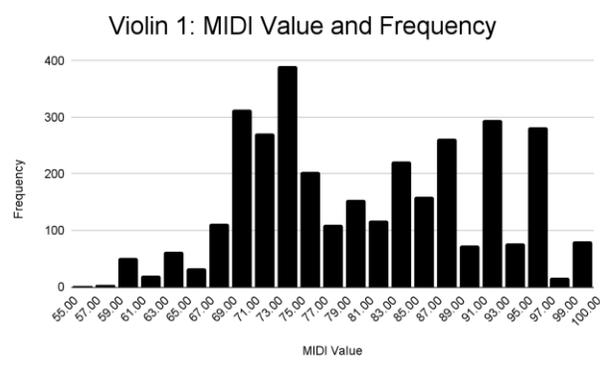
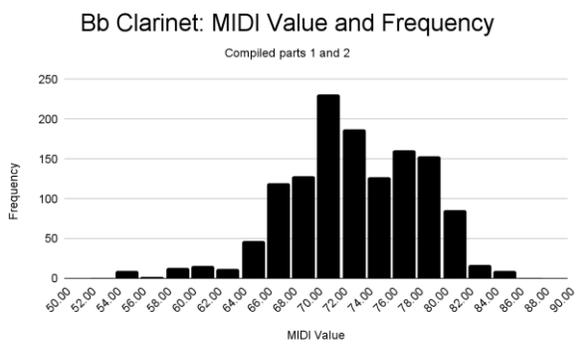
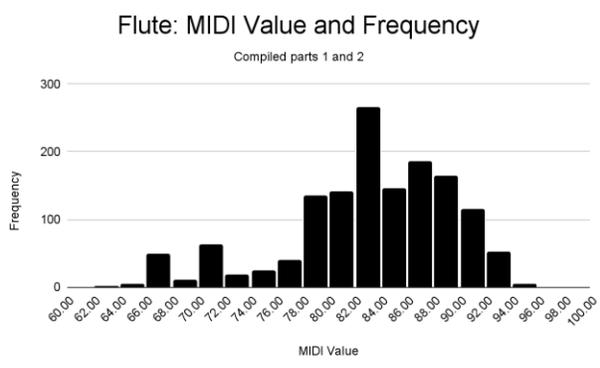
Eight instruments make up the bulk of this study. The two woodwind instruments include the Flute and the B flat Clarinet. The flute part includes both Flute 1 and Flute 2 parts within the same data set. The B flat Clarinet also includes both Clarinet 1 and Clarinet 2 parts within the same data set. The string instruments studied include the Violin, in which only the Violin 1 part will be included, and the Cello. Finally, brass instruments include the trumpet, trombone, and tuba. The percussion instrument included in the study is the Timpani, typically consisting of four notes at a time. Flute, Clarinet, Trumpet, and Violin are the high voices, typically given the melody. Trombone, Tuba, Cello, and Timpani have the bass lines, making them the low voices that form the root of the sound. The difference between the high voices and low voices may prove significant in the study when analyzing range, skewness, standard deviation, and normality.

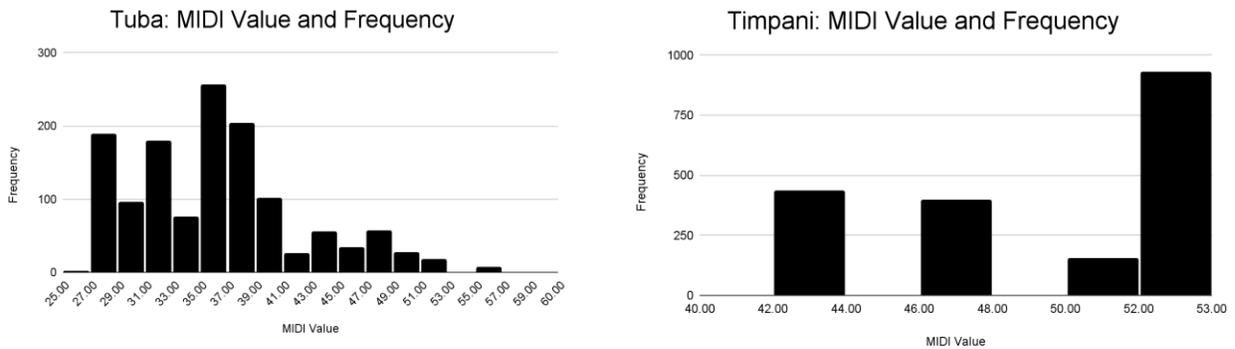
## **2. Main Body**

### **2.1 Collecting and Modelling Data**

Histograms were created to visually display the relationship between MIDI values and their frequency of occurrence for each instrument. First, MIDI files for each instrument were downloaded from MuseScore. A MIDI-to-CSV website created a dataset of note pitches as MIDI values, with their respective note name (i.e MIDI value: 70, note name: A#4). Data was copied into a Google Sheets spreadsheet, organized, and unused data, such as cue notes and grace notes, was deleted. A histogram of MIDI values was created, and descriptive statistics were computed for each instrument.

### **2.1.1 Histograms**





**Figure 1: Histograms**

### Histogram interpretations

The timpani part has three peaks and only a few bars, reflecting the limited number of notes the timpani can play at a time. The flute part appears to be left-skewed, with most of the data concentrated at high MIDI values, corresponding to higher pitches. This shows how the flute acts as a high voice in an orchestra. It is also unimodal, with a peak at MIDI value 83 (note: B5). The clarinet part is roughly symmetric and unimodal, with a peak at MIDI value 71 (note: B4). The violin 1 part has the widest range, with a peak at MIDI value 74 (note D5). The violin and clarinet have some of the largest ranges of playable notes among the instruments, making it reasonable to expect the data to be roughly symmetric for both. The cello part appears to be skewed right, reflecting its low voice, making up the low voices, and rarely playing high-pitched notes. The trumpet part is roughly symmetric and has a peak at MIDI value 74 (note: D5). The trombone part is roughly symmetric, with a peak at 60 (note: C4). Tuba is skewed right, also representing its low voice and its rare use of high-pitched notes.

### 2.1.2 Descriptive Statistics:

The dispersion and central tendency for each instrument were calculated according to the formulas shown and are summarized in the table 1 below.

Kurtosis describes the degree of “tailedness” in a probability distribution model. As a statistical measure, kurtosis indicates how concentrated the data are in the tails or peaks relative to a normal distribution. The three types of kurtosis are leptokurtic (kurtosis > 3), characterized by a high peak and heavy tails, mesokurtic (kurtosis = 3), similar to a normal distribution, and platykurtic (kurtosis < 3).

Formula for Kurtosis:

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^4$$

Formula for Skewness:

$$\frac{1}{n} \sum_{i=1}^n \left( \frac{x_i - \bar{x}}{s} \right)^3$$

**Table 1: Descriptive Statistics Evaluated from Histograms and Raw Data**

<b>Instrument</b>	<b>Mean</b>	<b>Median</b>	<b>Standard Deviation</b>	<b>Range</b>	<b>Kurtosis</b>	<b>Skewness</b>
Timpani	48.8152	51	3.6903	9	-10.8803	-0.5910
Flute 1	82.7041	83	6.1710	33	-9.9928	-0.8416
Bb Clarinet	72.5344	72	5.2408	34	-6.9742	-0.4244
Violin 1	80.3727	79	10.0130	43	32.0448	0.0636
Cello	49.6161	50	6.3171	31	-22.3233	0.3506
F Trumpet 1	69.7222	70	3.8897	29	1.7897	-0.7017
Trombone	57.9180	59	4.0608	20	-0.6799	-0.1457
Tuba	35.7024	35	6.1280	30	-6.8327	0.7725

**Kurtosis Interpretations**

The majority are platykurtic distributions, such as the timpani (-10.88), flute (-9.99), B-flat clarinet (-6.97), cello (-22.32), trombone (-0.68), and tuba (-6.83) parts, all showing negative kurtosis values. This shows that these instruments have a broad pitch range, rather than concentrating notes around a single pitch. However, the violin 1 plot is strongly leptokurtic, indicating a sharp peak. This shows that the violin part frequently centers its

notes around a specific pitch. This is consistent with the violin's role, largely performing the melody. The trumpet part shows a kurtosis value of 1.79, indicating slightly mesokurtic behavior. This represents its moderate pitch concentration, showing it plays an evenly distributed pitch register. This can be a mathematical explanation of the trumpet's role in playing both melodic and harmonic lines.

The descriptive statistical data will also be used for the correlation analysis

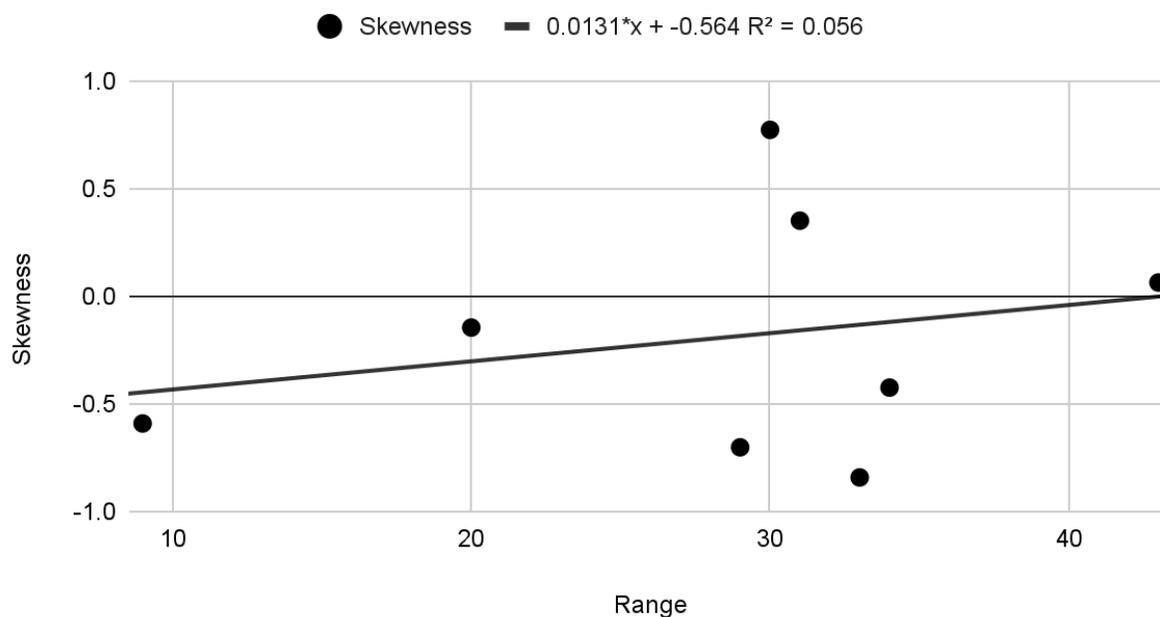
## **2.2 Correlation Analysis**

Descriptive statistics using histograms allowed us to calculate the individual pitch distributions for each instrument, but a correlation analysis is necessary to determine relationships between characteristics across instruments. A Pearson's correlation coefficient,  $r$ , will be calculated to examine if the pitch range of an instrument is mathematically related to skewness and kurtosis. By conducting this statistical analysis, it is possible to determine whether pitch range can be statistically related to symmetry and peakedness, or whether these features of the symphony are entirely governed by the instrumental roles within the orchestra.

### **2.2.1 Skewness vs. Range**

The correlation coefficient ( $r$ ) was calculated to determine whether skewness in values is correlated with an instrument's pitch range.

## Skewness vs. Range



$$r = 0.2375$$

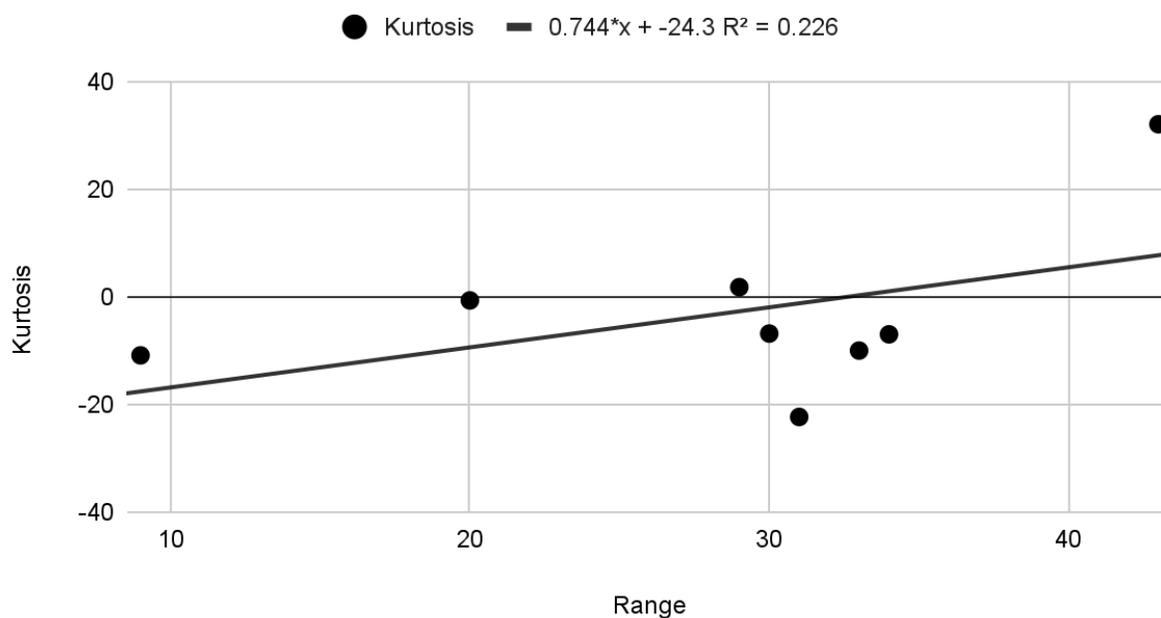
**Figure 2: Skewness vs. Range**

The correlation coefficient of 0.2375 indicates a weak positive relationship between range and skewness (Figure 2). This suggests that pitch range and skewness are weakly related, indicating that range itself does not correlate with symmetry. This relationship is not strong enough to conclude that pitch range could be a factor in determining the symmetry of a pitch distribution.

### 2.2.2 Kurtosis vs. Range

The correlation coefficient ( $r$ ) was calculated to determine whether the kurtosis of the values is correlated with an instrument's pitch range.

## Kurtosis vs. Range



$$r = 0.4750$$

**Figure 3: Kurtosis vs. Range**

There is a moderate positive relationship between kurtosis and range (Figure 3), implying that instruments with broader ranges tend to have a larger concentration of notes around particular pitch centers. However, the Violin 1 part's high kurtosis is an influential point in the dataset (the highest point on the right). Without it, the correlation coefficient would be  $-0.0905$ , indicating a weak relationship between kurtosis and range.

### **2.3 ANOVA Test**

ANOVA (Analysis of Variance) was used to determine whether there is a significant difference in mean pitch values across the three instrument families in the study.

Woodwind instruments include the flute and the B-flat clarinet. String instruments include the violin and the cello. Brass instruments include the trumpet, trombone, and tuba.

First conditions will be checked. Then, MIDI values for each instrument group will be compiled. ANOVA single-factor computation in Microsoft Excel will then be used to compute the ANOVA table terms. The information evaluated include SS (sum of squares), df (degrees of freedom), MS (mean square: the average variance calculated as the sum of squares divided by the degrees of freedom), F (F-Statistic: compares the two variances), P-value (evaluates the probability of obtaining an F-value this large by chance, assuming the null hypothesis is true), F crit (Critical F-Value: the threshold value from the F-distribution at a pre-selected significance level. In this study, a significance level of  $\alpha = 0.05$  will be used. A large F-value indicates that not all instrument pitch means are the same. The F-value and critical F-value will be compared. If  $F > F\text{-crit}$ , then the null hypothesis is rejected.

### **Conditions**

The condition of independence is not met, since the composer intentionally selects notes and instrumentation across a piece, implying that note value selections are influenced by other note values.

The normality condition will be evaluated by determining the skewness of each group.

### **Table 2: Skewness of Each Instrument Group**

	Woodwind	String	Brass
Skewness	-0.1347	-0.0686	0.1799

Each skewness value is approximately normal, the normality condition is met.

A homogeneity of variances will be evaluated by determining standard deviations of each group

**Table 3: Standard Deviation of Each Instrument Group**

	Woodwind	String	Brass
Standard Deviation	9.6674	17.2916	15.8565

The largest standard deviation, 17.2916, is less than the two times the smallest, 9.6674.

This assumption is met.

Since independence is not met, we will proceed with caution.

### **Hypothesis**

Null Hypothesis: The mean pitch is the same for woodwinds, strings, and brass

Alternative hypothesis: At least one instrument family has a different mean pitch

### **Calculations**

**Table 4: Summary for Instrument Groups**

	Count	Average	Variance
Woodwinds	2744	77.8083	58.7886
Strings	5288	68.9030	298.9981
Brass	2498	49.1853	251.4286

**Table 5: ANOVA Result**

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	1130302.57	2	565151.28	2510.40	0	2.9966
Within Groups	2369877.59	10527	225.1237			
Total	3500180.16	10529				

### Conclusion

The calculated F-value of 2510.40 is far greater than the critical value of 2.9966. The p-value rounds to 0, far below the significant level of  $\alpha=0.05$ . Therefore, the null hypothesis that all instrument groups have equal mean pitch is rejected.

### **3. Final Conclusion:**

In conclusion, this investigation successfully evaluated a comparison of pitch distributions between different instruments in Dvořák's Symphony No. 9, Movement IV. It evaluates the shapes of these distributions in terms of skewness, modality, and kurtosis, and explains how these characteristics relate to the instrument's role in a symphony. Additionally, results indicate that at least one instrument group's mean pitch is significantly different from other groups, suggesting that instrument type is a major contributor to pitch variation.

Descriptive Statistics showed that melodic instruments such as the flute, violin, and clarinet exhibit wide ranges, reflecting their flexibility in register. The flute, in particular, is left-skewed, likely because the composer chose ranges of notes more suitable to the instrument, which explains why the majority of the flute's MIDI values represent higher pitches. Harmonic instruments, such as the tuba and cello, were right-skewed, consistent with their role as lower voices in symphonies, bringing out the symphonic texture. Kurtosis calculations indicate that the majority of instruments exhibit a broad range rather than concentrated notes around a peak. These descriptive statistics effectively quantitatively determines how Dvořák's composition choices were shaped by the pitch range of instruments.

Correlational analyses confirmed that the correlation between range and skewness is weak, but that between range and kurtosis is stronger. This showed that instruments with larger ranges may exhibit more pronounced pitch clustering, which closely aligns with my own observations as an orchestral musician.

ANOVA test rejected the null hypothesis, proving that differences in pitch instruments observed were unlikely to be caused by systematic differences in orchestral roles

However, limitations are present throughout my study. Furthermore, MIDI pitch values were gathered from online sheet music; as a result, specific frequency values for each note could not be collected, and the data were treated discretely. This is not accurate to the real world, however, since musicians typically do not always perform repertoire in tune. This may have created small distortions in kurtosis calculations and descriptive statistical results. Additionally, the scope of this study is highly limited, evaluating only a few instruments and examining only the most famous movement of the symphony.

Furthermore, since the independence assumption failed, the results are not completely reliable. Results are only approximations. This may have led to an inflated F-value, increasing the risk of a Type I error. As a result, the test may claim a significant difference even though the effect may be exaggerated. In the scope of the test, however, the results and conclusions drawn from the study are reasonable, since composers likely do not use significant amounts of chance variation in their compositional works.

Further exploration could include accurately collecting MIDI values from a live performance rather than from sheet music. This will make results more applicable to the real world.

This exploration demonstrated that statistics can effectively be used to model and explain musical structure, revealing how pitch distributions are determined by instrument ranges and compositional decisions. Overall, Dvořák's use of pitch is intentional, not based on chance, and illustrates an intersection between composition and mathematical reasoning.

## Works Cited

*MIDI Converter*, <https://midi-to-csv.vercel.app/>. Accessed 23 October 2025.

*MuseScore: Free music composition and notation software*, <http://musescore.org>.  
Accessed 23 October 2025.

Smith's, Dave, and Peter Howell. "MIDI - Wikipedia." *Wikipedia, the free encyclopedia*, <https://en.wikipedia.org/wiki/MIDI>. Accessed 14 December 2025.