Quantifying Eurovision success: using regression to predict song contest scoring.

Racquel Pickett & Yawo Zewu

Group 11

December 12, 2025

STAT 311 Regression Analysis

Fall 2025

Quantifying Eurovision success: using regression to predict song contest scoring.

Introduction

The Eurovision Song Contest is an annual song competition consisting of participants primarily from the European region. Each contestant represents their country with approximately 42 countries represented each year. The competition is held near May and consists of semi-final and final performances. Each year's winner is expected to host the next year's competition (Eurovision Song Contest, 2025).

Rules for the competition involve having an original song less than three minutes long, lead vocals must be live, and a maximum of six performers are allowed on the stage. Performances receive scores from a jury representing each country and a popular vote. These are tallied into semi-final and grand final scores (Eurovision Song Contest, 2025).

This project aimed to answer several research questions.

1. Do songs performed in minor key receive a higher final score than those performed in major key?
2. Is a higher semi-final score associated with a higher final score?
3. Do songs performed in English receive higher final scores than those performed in other languages?
4. Does type of emotional element influence final score?
5. Is song duration related to final score?

Understanding how these variables contribute to a higher score can help participants identify influential characteristics that they may consider building into their song created for Eurovision. The research group aimed to study these questions by using various descriptive and regression analysis techniques.

Data Description

Data for this project includes a GitHub Excel file obtained from six Eurovision contests (2017-2024). Each year Eurovision results are made publicly available and posted on various sites for consumer consumption. There were 45 unique countries represented with 180 total rows of data. Both qualitative and quantitative variables were in the file consisting of song characteristics, scoring and country representation.

Dependent Variables: Final Scoring

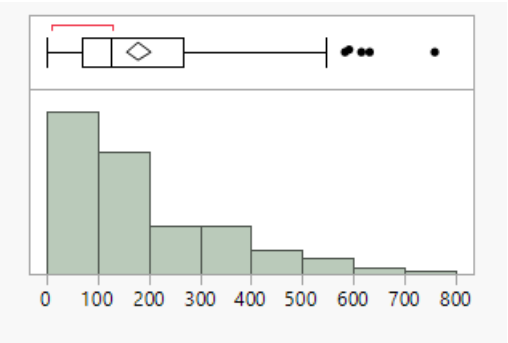
Independent Variables:

• Semi-Final Score • Language • Beats Per Minute • Danceability • Chords	• Length • Sadness • Joy • Love	• Anger • Fear • Surprise • Country

Data was examined to determine if any cleaning or modifications were needed prior to running descriptive analyses. Language was recoded to group English versus Other language of performance. Country was excluded due to the complexity of having 45 countries represented in a categorical variable. Data missing any semi-final scores were excluded from the analysis. This occurred when a country was allowed into finals without being required to compete in other stages of the process.

Exploratory Data Analysis

A histogram chart was selected to display the distribution across semi-final scoring. As can be seen, song scoring had a right-skewed distribution indicating fewer songs had a high score. This was expected since each jury member only provides scores to their favorite 10 out of 40+ performers.



Quantiles

100.0%	maximum	758
99.5%		758
97.5%		599.8
90.0%		375.3
75.0%	quartile	258.25
50.0%	median	128
25.0%	quartile	78
10.0%		49.6
2.5%		25.425
0.5%		16
0.0%	minimum	16

Summary Statistics

Mean	184.15942
Std Dev	146.52711
Std Err Mean	12.473216
Upper 95% Mean	208.82435
Lower 95% Mean	159.49449
N	138
N Missing	0

Figure 1

Graphical representation of the emotions in songs showed some variability. Datapoints represent a proportion of how much emotion was in each song, where a value of zero meant the emotion was not identified. Sadness, joy and anger were similarly distributed with a mix of being included songs. Love, fear and surprise were used less frequently indicated by more dots being on the zero line. When love was used, it appeared to have a stronger relationship to final

points versus other emotions. Since non-parallel lines were noted, it could be valuable to look for interactions.

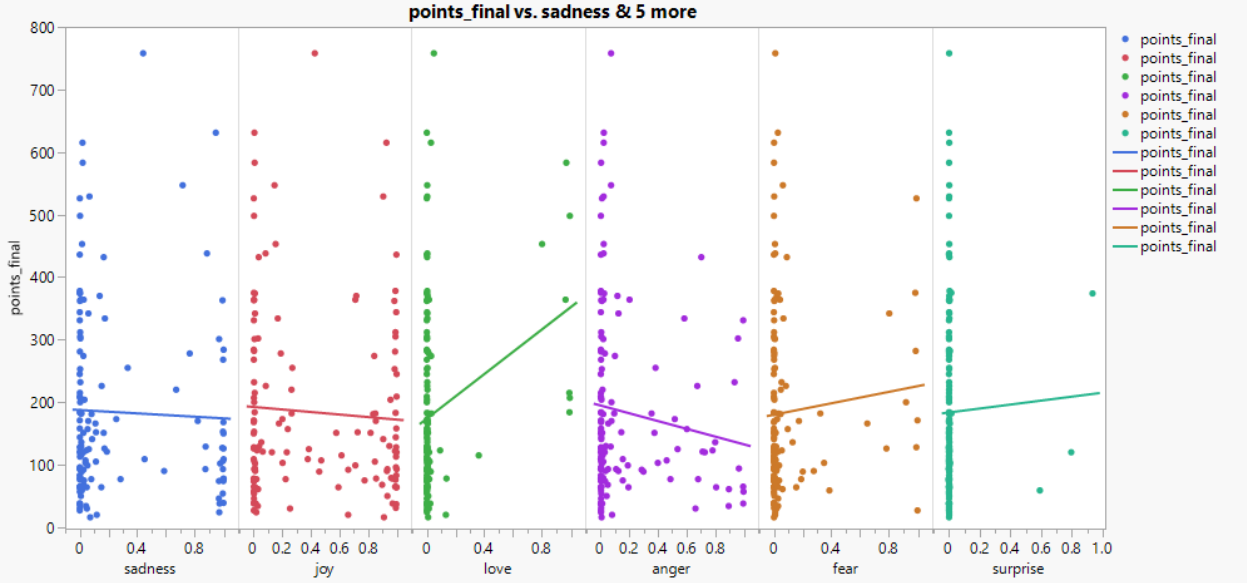


Figure 2

Language of song and chord played were similarly distributed with both implying no relationship to final scoring. Length of song was typically between 150 and 225 seconds with one outlier above 300. Danceability had a similar more spread-out pattern. Semi-final scoring appeared to have a strong positive relationship with final scoring.

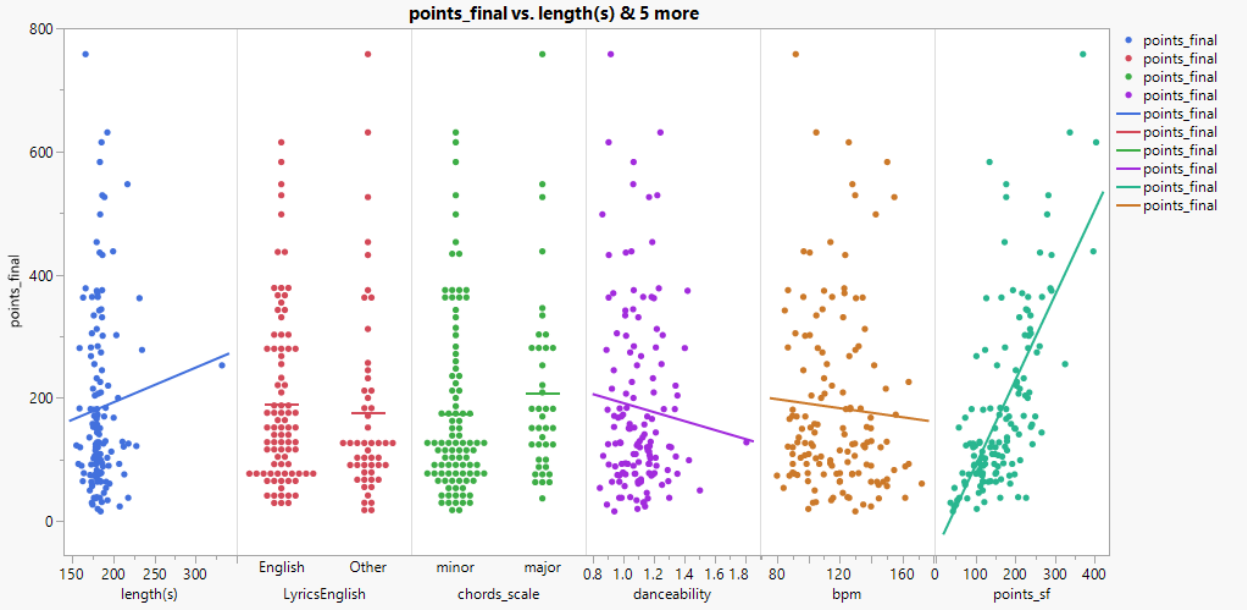


Figure 3

Correlation data supported a high positive relationship between final and semi-final points. Love showed a low positive relationship. All other independent variables had much smaller correlations to the dependent variable. When looking between independent variables, there were notable relationships such as: fear-sadness, fear-joy, bpm-semi-final, bpm-fear, danceability-surprise, anger-sadness, joy-anger, and danceability-sadness. These correlations suggest the need for a multiple versus simple regression analysis.

Correlations

	points_final	points_sf	sadness	joy	love	anger	fear	surprise	length(s)	bpm	danceability
points_final	1.0000	0.6869	-0.0352	-0.0591	0.2581	-0.1247	0.0751	0.0249	0.0713	-0.0526	-0.0728
points_sf	0.6869	1.0000	-0.0524	0.0282	0.1002	-0.0410	-0.0232	0.0399	-0.0715	-0.1779	0.0288
sadness	-0.0352	-0.0524	1.0000	-0.5216	-0.1660	-0.2506	-0.1797	-0.0963	0.0708	-0.0573	-0.1465
joy	-0.0591	0.0282	-0.5216	1.0000	-0.1925	-0.3337	-0.3023	-0.1260	-0.0218	0.0669	0.1270
love	0.2581	0.1002	-0.1660	-0.1925	1.0000	-0.1299	-0.0967	-0.0358	-0.0434	0.1056	-0.1175
anger	-0.1247	-0.0410	-0.2506	-0.3337	-0.1299	1.0000	-0.0796	-0.0733	-0.0242	0.0641	-0.0630
fear	0.0751	-0.0232	-0.1797	-0.3023	-0.0967	-0.0796	1.0000	0.0222	0.0181	-0.1778	0.1010
surprise	0.0249	0.0399	-0.0963	-0.1260	-0.0358	-0.0733	0.0222	1.0000	-0.0528	-0.0441	0.1870
length(s)	0.0713	-0.0715	0.0708	-0.0218	-0.0434	-0.0242	0.0181	-0.0528	1.0000	0.1040	-0.0390
bpm	-0.0526	-0.1779	-0.0573	0.0669	0.1056	0.0641	-0.1778	-0.0441	0.1040	1.0000	-0.1230
danceability	-0.0728	0.0288	-0.1465	0.1270	-0.1175	-0.0630	0.1010	0.1870	-0.0390	-0.1230	1.0000

The correlations are estimated by Row-wise method.

Figure 5

Methods & Results

Based on the distribution charts, a linear first-order multiple regression model was used to test whether there were significant relationships between independent variables and the dependent variable final score. Residual analysis, outlier correction, interaction considerations and step-wise regression were included to make a final model selection.

Model 1 First Order

$$\epsilon(y) = \mathrm{B}_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12}$$

$X_1 = SemiFinal$ $X_7 = Surprise$

$X_2 = Sadness$ $X_8 = Length$

$X_3 = Joy$ $X_9 = BPM$

$X_4 = Love$ $X_{10} = Danceability$

$X_5 = Anger$ $X_{11} = Chord$

$X_6 = Fear$ $X_{12} = English$

Model 1 Hypothesis

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0$$

H_a: *at least one of the* $\beta_j \neq 0$ at the 0.05 significance level

J=1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Model 1 Results

Model 1 showed an overall significance with a p-value of less than 0.001. In the parameter estimates section there were biased results listed. This can occur when there is high multicollinearity, or when there are not enough observations to account for the number of variables in the model. Examining the data suggested the variable surprise had a small number of observations that were not zero. Removing the variable appeared to be the best method for moving forward.

Summary of Fit

RSquare	0.523178
RSquare Adj	0.481878
Root Mean Square Error	107.984
Mean of Response	187.0863
Observations (or Sum Wgts)	139

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	1624856.9	147714	12.6679
Error	127	1480888.1	11661	Prob > F
C. Total	138	3105745.0		<.0001*

Parameter Estimates

Term		Estimate	Std Error	t Ratio	Prob>|t|
Intercept	Biased	-171.8445	162.9456	-1.05	0.2936
points_sf		1.4273111	0.135776	10.51	<.0001*
LyricsEnglish[Other]	Biased	10.419704	10.35475	1.01	0.3162
sadness	Biased	-7.163183	84.32327	-0.08	0.9324
joy	Biased	-35.88262	82.85785	-0.43	0.6657
love	Biased	89.048165	92.85895	0.96	0.3394
anger	Biased	-64.73659	86.33621	-0.75	0.4548
fear	Biased	36.178479	91.98203	0.39	0.6947
surprise	Zeroed	0	0	.	.
length(s)		0.8586588	0.509114	1.69	0.0941
bpm		0.6894856	0.456033	1.51	0.1330
danceability		-90.26132	66.81952	-1.35	0.1792
chords_scale[major]		-2.818121	10.41146	-0.27	0.7871

Figure 6

Model 2 First Order Surprise Removed

$$\epsilon(y) = \mathrm{B}_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12}$$

$X_1 = SemiFinal$ $X_8 = Length$

$X_2 = Sadness$ $X_9 = BPM$

$X_3 = Joy$ $X_{10} = Danceability$

$X_4 = Love$ $X_{11} = Chord$

$X_5 = Anger$ $X_{12} = English$

$$X_6 = Fear$$

Model 2 Hypothesis

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = 0$$

$$H_a: at\ least\ one\ of\ the\ \beta_j \neq 0 \text{ at the 0.05 significance level}$$

J=1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12

Model 2 Results

Model 2 showed an overall significant relationship with a p-value of less than 0.0001. In the indicator parameterization, semi-final points showed a significant positive relationship with final points. No other independent variables were showing a significant contribution to the model.

Summary of Fit

RSquare	0.523178
RSquare Adj	0.481878
Root Mean Square Error	107.984
Mean of Response	187.0863
Observations (or Sum Wgts)	139

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	1624856.9	147714	12.6679
Error	127	1480888.1	11661	Prob > F
C. Total	138	3105745.0		<.0001*

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-185.0823	162.5391	-1.14	0.2570
points_sf	1.4273111	0.135776	10.51	<.0001*
LyricsEnglish[Other]	20.839407	20.7095	1.01	0.3162
sadness	-7.163183	84.32327	-0.08	0.9324
joy	-35.88262	82.85785	-0.43	0.6657
love	89.048165	92.85895	0.96	0.3394
anger	-64.73659	86.33621	-0.75	0.4548
fear	36.178479	91.98203	0.39	0.6947
length(s)	0.8586588	0.509114	1.69	0.0941
bpm	0.6894856	0.456033	1.51	0.1330
danceability	-90.26132	66.81952	-1.35	0.1792
chords_scale[minor]	5.6362427	20.82292	0.27	0.7871

Figure 7

Studentized residuals showed row 154 was outside the limits for an outlier. There were other outliers between the green and red lines, but row 154 was most notable. Normal QQ Plot was close to the predicted line with row 154 in the upper righthand corner.

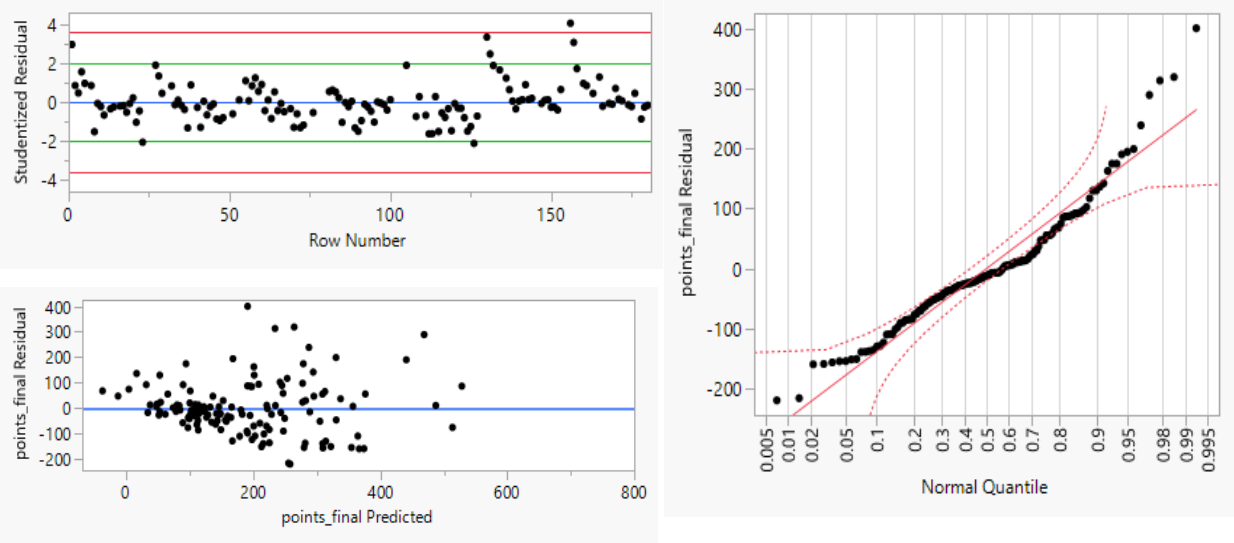


Figure 8

Cook's D was at a 0.35 and leverage was at 0.08, indicating the datapoint was not necessarily influencing the data. Checking to see what would change when removing row 156, resulted in a slightly better adjusted R square values without compromising overall significance or indicator parameterization values. Residual Plot and Normal QQ Plot remained nearly identical.

Summary of Fit

RSquare	0.555467
RSquare Adj	0.516658
Root Mean Square Error	101.8697
Mean of Response	184.1594
Observations (or Sum Wgts)	138

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	11	1633859.3	148533	14.3130
Error	126	1307557.2	10377	Prob > F
C. Total	137	2941416.5		<.0001*

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-148.9933	153.5899	-0.97	0.3339
points_sf	1.4269653	0.128089	11.14	<.0001*
LyricsEnglish[Other]	25.745873	19.57374	1.32	0.1908
sadness	-19.66185	79.60749	-0.25	0.8053
joy	-34.26294	78.16729	-0.44	0.6619
love	93.800295	87.60882	1.07	0.2864
anger	-60.24405	81.45512	-0.74	0.4609
fear	31.799439	86.78045	0.37	0.7147
length(s)	0.8264781	0.480352	1.72	0.0878
bpm	0.3645536	0.437497	0.83	0.4063
danceability	-81.67466	63.07108	-1.29	0.1977
chords_scale[minor]	-0.110823	19.69416	-0.01	0.9955

Figure 9

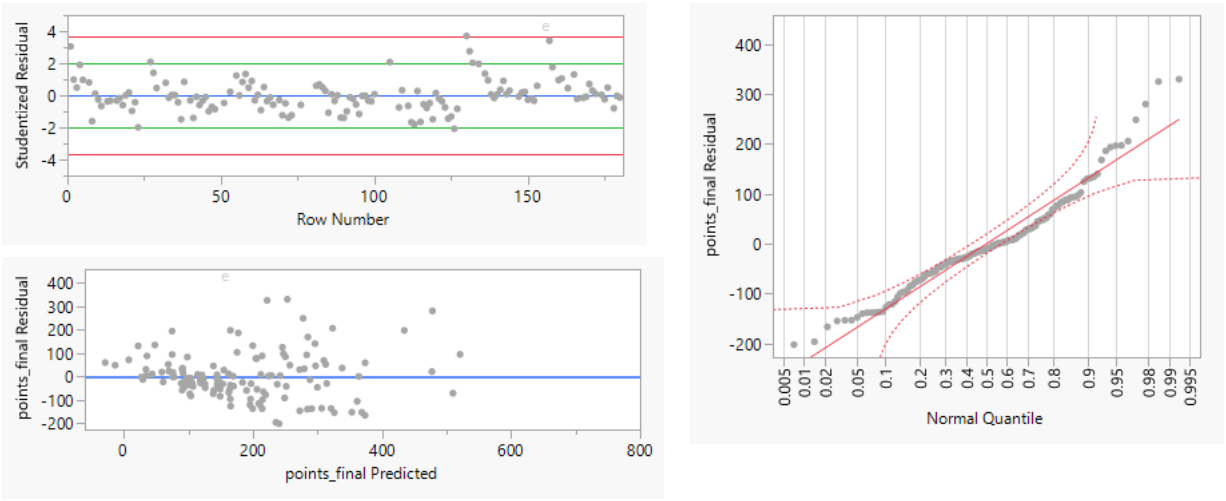


Figure 10

Model 3 Stepwise All Models

Stepwise regression was used to identify the best model depending on how many independent variables were included in the analysis. Figure 11 shows the number of variables in the model across the x-axis and the values being compared for identifying the best model along the y-axis. A maximum of 10 variables was used for this graph. The best model should have the highest R square, lowest RMSE and lowest Cp value. R square can be seen increasing as the number of independent variables being included goes up with 6-10 being highest. However, Cp values are lowest between four and six variables. RMSE is lowest with six through eight variables. BIC was lowest between three to four variables. Based on this information, four variables seem to be the best option for choosing a model. The variables included in that model were: semi-final points, performance language, love, fear, length and danceability.

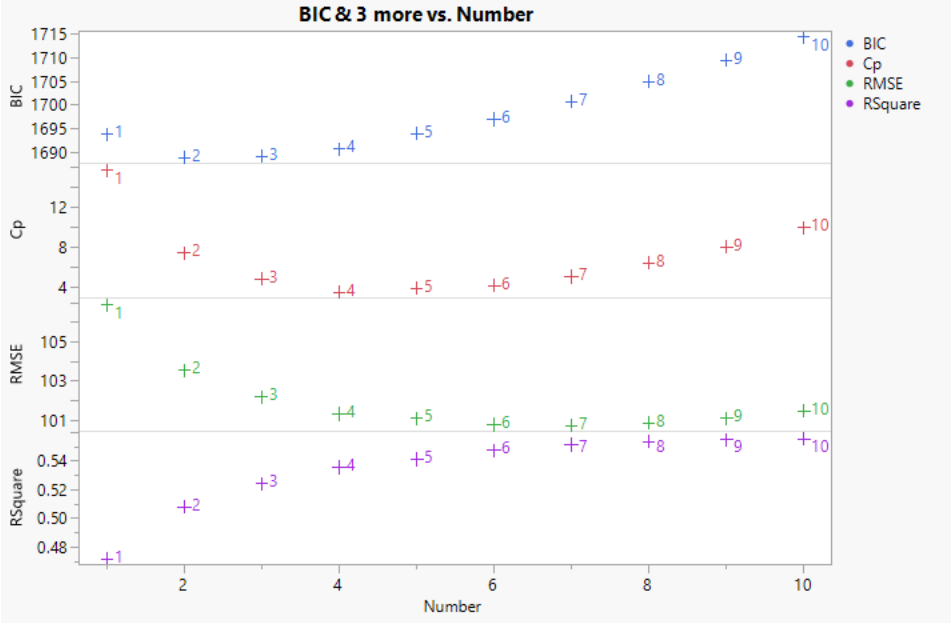


Figure 11

Testing this model included the following variables and hypothesis.

$$\epsilon(y) = \mathrm{B}_0 + \beta_1 X_1 + \beta_4 X_4 + \beta_6 X_6 + \beta_8 X_8$$

$$X_1 = SemiFinal$$

$$X_4 = Love$$

$$X_6 = Fear$$

$$X_8 = Length$$

Model 3 Hypothesis

$$H_0: \beta_1 = \beta_4 = \beta_6 = \beta_8 = 0$$

H_a: *at least one of the* $\beta_j \neq 0$ at the 0.05 significance level

J=1, 4, 6, 8

Model 3 Results

This model had a moderate adjusted R squared value and a low significant p-value. Semi-final points, love, and length all showed significant betas indicating all three were contributing to the model. Even though fear was not significant, it was close to 0.05.

Studentized residuals, residual plots and QQ plot could use some improvements, but were generally good considering only having a moderate significant relationship.

Summary of Fit

RSquare	0.535913
RSquare Adj	0.521956
Root Mean Square Error	101.3099
Mean of Response	184.1594
Observations (or Sum Wgts)	138

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	4	1576344.3	394086	38.3961
Error	133	1365072.2	10264	Prob > F
C. Total	137	2941416.5		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-243.7718	90.61736	-2.69	0.0081*
points_sf	1.3643957	0.119797	11.39	<.0001*
love	142.51531	41.25739	3.45	0.0007*
fear	66.216674	36.24222	1.83	0.0699
length(s)	1.0019118	0.468768	2.14	0.0344*

Figure 12

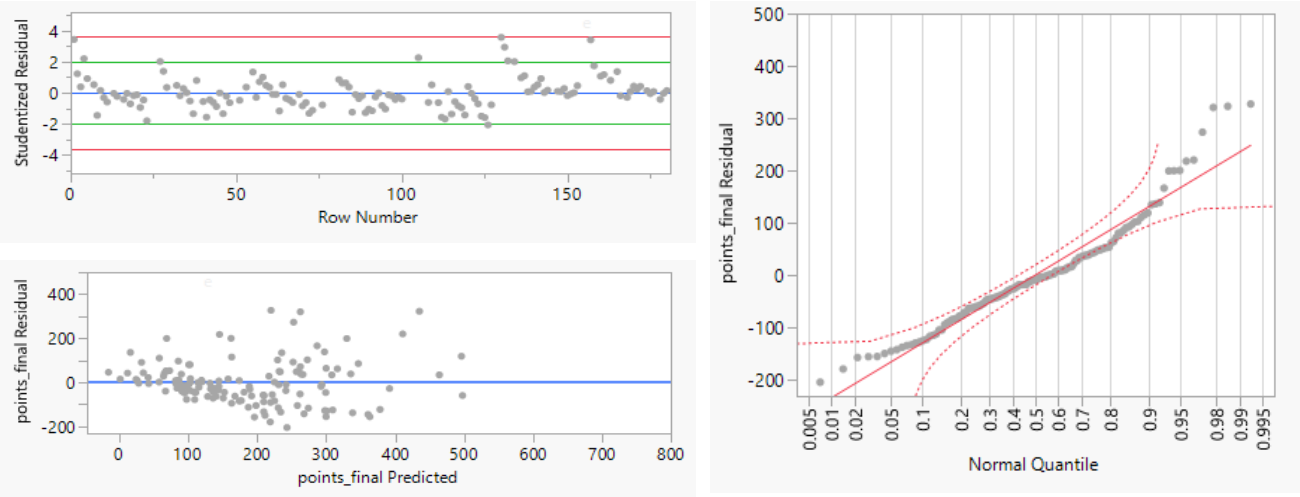


Figure 13

Model 4 First Order with Country Added & Semi-Final Score Removed

Model 4 examined the removal of semi-final score while adding country represented by the performer to see if country had any effect on final score. Biased estimates were seen again with the inclusion of Surprise, even though semi-final was not part of the model. This further suggests the issue was related to Surprise versus multicollinearity. Due to the result, surprise was removed for this model.

$$\epsilon(y) = \mathrm{B}_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} + \beta_{11} X_{11} + \beta_{12} X_{12}$$

$X_1 = SemiFinal$ $X_9 = BPM$

$X_2 = Sadness$ $X_{10} = Danceability$

$X_3 = Joy$ $X_{11} = Chord$

$X_4 = Love$ $X_{12} = English$

$X_5 = Anger$ $X_{13} =$ Eastern Caucasus Other

$X_6 = Fear$ $X_{14} = Southern\ Balkan\ Europe$

$X_8 = Length$

Model 4 Hypothesis

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_8 = \beta_9 = \beta_{10} = \beta_{11} = \beta_{12} = \beta_{13} = \beta_{14} = 0$$

$H_a: at\ least\ one\ of\ the\ \beta_j \neq 0$ at the 0.05 significance level

J=1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14

Model 4 Results

This model had a very small adjusted R squared value and was overall not-significant with a p-value greater than 0.05. None of the betas in the indicator parameterization were significant and the Normal QQ Plot showed deviations outside of the expected dotted line bounds. Fail to reject the null.

Summary of Fit	
RSquare	0.134152
RSquare Adj	0.044104
Root Mean Square Error	146.6725
Mean of Response	187.0863
Observations (or Sum Wgts)	139

Analysis of Variance				
Source	DF	Sum of Squares	Mean Square	F Ratio
Model	13	416643.2	32049.5	1.4898
Error	125	2689101.8	21512.8	Prob > F
C. Total	138	3105745.0		0.1301

Indicator Function Parameterization

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	632.50039	355.4056	1.78	0.0776
Cntry3Reg[Eastern/Caucasus/Other]	-5.16939	31.24647	-0.17	0.8689
Cntry3Reg[Southern/Balkan Europe]	-35.15705	31.42556	-1.12	0.2654
LyricsEnglish[English]	27.190971	28.51773	0.95	0.3422
sadness	-38.29824	115.9077	-0.33	0.7416
joy	-40.77839	113.9844	-0.36	0.7211
love	145.6032	128.1209	1.14	0.2579
anger	-84.07798	119.1156	-0.71	0.4816
fear	27.165286	126.9887	0.21	0.8310
length(s)	-2.020745	1.943123	-1.04	0.3004
bpm	-3.722393	2.384316	-1.56	0.1210
beats_count	1.3246336	0.868026	1.53	0.1295
danceability	-64.41877	91.65504	-0.70	0.4835
chords_scale[minor]	-23.85724	28.13327	-0.85	0.3981

Figure 14

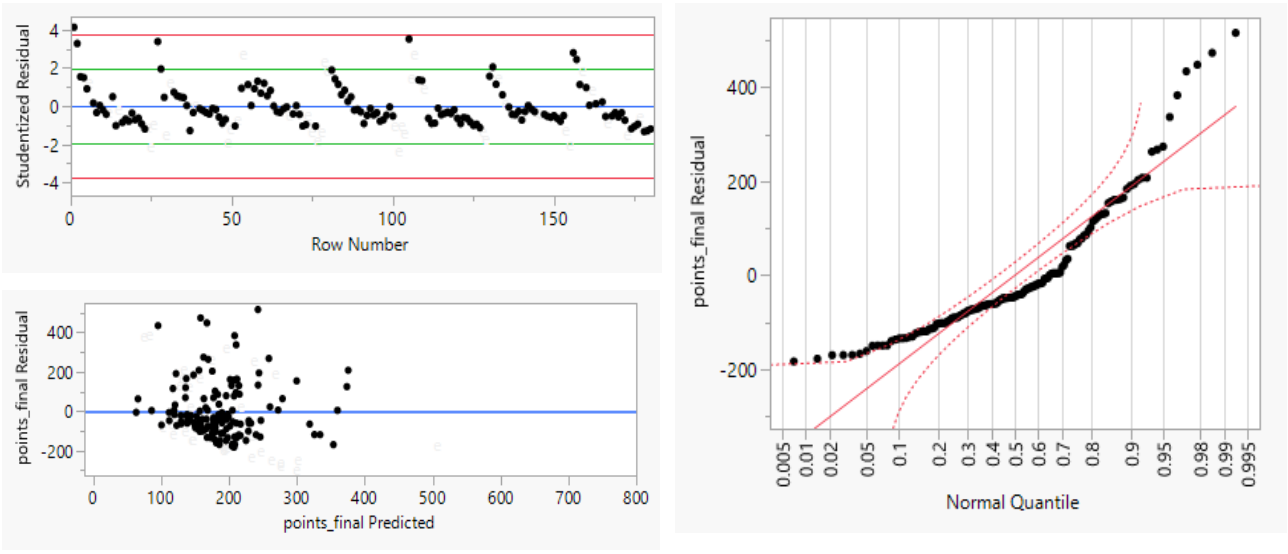


Figure 15

Model Comparison

The table below compares the four models across adjusted R squared, mean square error, p-value, number of independent variables, and number of observations. An ideal model would have a moderate to high adjusted R square, low p-values and a manageable number of independent variables. Model 1 had issues with the bias error due to an inability to compute formulas from the small number of observations with the Surprise element. After removing this variable, model 2 showed some improvements in R squared with a reduction in MSE. Model 4 added in country as a variable, but failed to show any type of relationship to the dependent variable final scoring. Model 3 had the smallest number of independent variables, highest R squared value and lowest MSE for the four options. Thus, Model 3 is the best option across these measures and allows for the most simplistic formula with the most impact.

	Adj R^2	MSE	P-value	N	# IV's	Features
Model 1	0.48	107.98	<0.0001	139	12	Bias Error
Model 2	0.51	101.86	<0.0001	138	11	FirstOder-Surprise-Outlier
Model 3	0.52	101.30	<0.0001	138	4	Stepwise All Models
Model 4	0.04	146.67	0.1301	139	13	FirstOrder+Cntry-SemiFinal

Conclusion & Discussion

The purpose of this analysis was to identify key factors that contribute to a higher final score in the Eurovision Song Contest. Understanding these relationships could provide useful information to prospective performers when designing songs to influence musical structure, or other elements. The dataset included variables such as country being represented, the chord a song was played, various emotional attributes, language of song, length, beats per minute and danceability. Initially there were four questions were to be answered through this analysis.

1. Do songs performed in minor key receive a higher final score than those performed in major key?
2. Is a higher semi-final score associated with a higher final score?
3. Do songs performed in English receive higher final scores than those performed in other languages?
4. Does type of emotional element influence final score?
5. Is song duration related to final score?

Depending on which model was selected to answer a question, the results would vary. Speaking from the best model selected only, songs performed in minor key did not receive a higher score than those performed in major key. Higher semi-final scores were associated with a higher final score. Songs performed in English were not more likely to receive a higher score versus those performed in other languages. Type of emotional attribute did influence the final score when love or fear were included. Song duration was related to final score where longer songs had a higher score. Future analyses could focus on determining additional variables to explain a higher proportion of the variance.

Appendix

1. Eurovision Data – Fit Least Squares Best Model 4 Variables Stepwise
2. EurovisionData – Distribution Histogram
3. EurovisionData – Fit Least Squares Best Model Stepwise Outcome
4. EurovisionData – Fit Least Squares Biased Surprise
5. EurovisionData – Fit Least Squares Exclude Outlier 156
6. EurovisionData – Fit Least Squares Interactions
7. EurovisionData – Fit Least Squares Row 156 Included
8. EurovisionData – Fit Stepwise Backward from All Interactions Input
9. EurovisionData – Fit Stepwise First Order
10. EurovisionData – Graph Builder Figure 1 Independent Variables
11. EurovisionData – Graph Builder Figure 2 Independent Variables
12. EurovisionData – Multivariate Correlations
13. EurovisionDataCountry (dataset)
14. EurovisionModelComparison – Graph Builder
15. EurovisionModelComparison (dataset)
16. EurovisionDataCountry – Fit Least Squares

References

Premarathna, I. (2025). Stat 311 Regression Analysis Fall 2025 [PowerPoint presentation]. D2L Brightspace.

Adamska, K. (2025). Eurovision Song Science. GitHub repository, https://github.com/kasiadamska/eurovision-song-science/blob/main/data

OpenAI. (2024). *ChatGPT* How to group Eurovision Song Contest participant countries into common regions. (GPT-5.1) [Large language model]. Retrieved December 1, 2025, from https://chat.openai.com/

Eurovision Song Contest. (December 2025). Eurovision.tv. https://eurovision.tv/