# Letter

**To**: the Puzzle Editor of the New York Times

**Date**: Monday, February 20 , 2023

**Subject**: Wordle's Development Forecasts and Related Recommendations

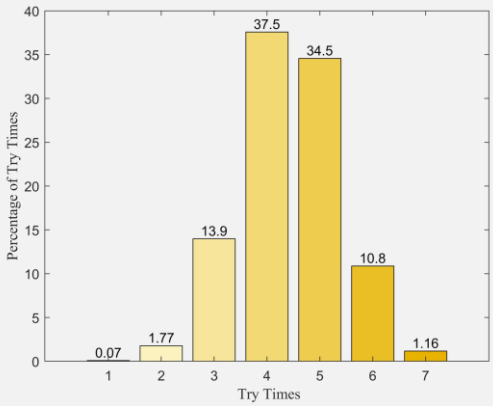
---

## Dear Mr. Editor :

With the help of the data provided by the MCM organizing committee, we analyze in detail the daily results of Wordle from January 7, 2022 to December 31, 2022. Wordle exploded thanks to its creative gameplay and the interesting puzzles provided by Puzzle Editor every day. While we are marveled at the success of wordle, we also find some interesting phenomena through the data set and make some guesses about the future development of wordle, which I would like to share with you next.

First of all, we develop a prediction model for the number of daily reported results for wordle, mainly based on the **ARIMA (1,1,1)** model. It is predicted that **14,869** reports will be reported on March 1, 2023, probably due to the decline in wordle's popularity over time. We also find through analysis of the data that the attributes of the words do not affect the percentage of players who choose the hard mode that day, which is understandable because no one knows the difficulty of the words before they start the game, and thus does not affect their mode choice.

Secondly, we develop a model for predicting the distribution of daily attempts, based on **coupled double discrete normal distribution** to fit the overall distribution. Using this model, we predict the distribution of attempts for the word EERIE on March 1 as shown on the right. The model is subject to many uncertainties such as **significant events, time lapse, word attributes**... and we can quantify our confidence in the model prediction in the form of **confidence intervals**.




Then, using **multiple linear regression**, we develop a difficulty classification model based on the various attributes of the words and determine a star classification model for difficulty through **boundary optimization**, concluding that **EERIE is a 4-star difficulty word** (the most difficult category of words).

Finally, we also find some interesting features of this dataset, such as

1) *The amount of percentage change in the number of people who choose hard mode on that day is negatively correlated with the difficulty coefficient of yesterday's word*
2) *The percentage of One-try guesses is negatively correlated with the number of repeated letter in the word*

We hope that our prediction and classification models and the interesting features we summarize will help New York Times to better develop wordle as an excellent daily game.

**Sincerely yours**