Wordle Predictive Analysis of Results

Summary

This article studies, predicts, and analyzes the results of daily reports on the puzzle game wordle. According to the existing conditions, based on ARIMA time series model, Kmeans clustering analysis, Shapiro-Wilk normality test and other methods to comprehensively consider different situations to forecast the results of the report, and establish relevant models to get the prediction of the percentage of people and the distribution of the results of the report. The difficulty of words is classified according to various possible attributes of words. Finally, some other interesting features of the data set are described.

In order to predict the number of results reported on March 1, 2023, we first preprocessed the original data from January 7, 2022 to December 31, 2022, corrected the outliers with a Hample filter, and found and corrected the outliers at 529. Then, after testing the original data with the ADF stationarity test, autocorrelation and partial autocorrelation analysis were carried out. The model was determined as ARIMA(3,0,0) model according to the truncated and trailed features in autocorrelation gragh and thus can be solved with the provided data. Finally, we test the model's residual with Ljung-Box white noise test, and it turns out that the residual is white noise, which means the model is reliable. We then use the model to predict that the number of results reported on March 1, 2023 will be 20,459, with a fluctuation from 18,829 to 22,289.

In order to predict the relevant percentage of the number of attempts (1, 2, 3, 4, 5, 6, x) on a certain date in the future, we try to sipecify and parameterize the distribution first, and define the parameter on the next step. First, the S-W method was used for normality test, and the obtained data all passed the normality test, so the normal distribution could be used to fit the data. Secondly, homogeneous variance is used to test whether different words will affect the variance. After passing the test, only the parameter μ is needed to be determined. By analyzing word frequency, number of letters, vowel and consonant combinations, and connection weirdness, we extract several critical parameters to describe the word's attribute. After parameterization, we use multiple linear regression to fit the extracted parameters to μ in provided data. The final result for the word Eerie, 0, 1, 6, 21, 34, 26, and 11, for 1~7 trise respectively.

In order to classify words according to difficulty, we used the parameters obtained before to construct the feature space, and applied Kmeans clustering method to classify words in this space. Five factors, including frequency of word use, number of letters appearing, combination of vowels and consonants, combination of letters, and connection weirdness, were used as parameters affecting the classification of a word. After cluster analysis, the words are divided into five categories, and the word Eerie belongs to the second category. With manually labelling, we can determine that category 2 is the hardest, followed with 4,1,5,3, which means 'eerie' could be one of the hardest wordle.

In addition, looking at the raw data, we found some other characteristics. For example, the percentage of people trying 1 and 2 is much smaller than the percentage of people trying other attempts. We try to explain this interesting phenomenon from the point of view of information entropy, and venture to guess that no algorithm of any kind can reduce the average number of guesses below 3.

Key words: ARIMA temporal model, parameterized distribution, multiple linear regression, K-means clustering analysis.

Contest

1Background and Restatement Of the Question

1.1 Question Background

Wordle was the first puzzle game to hit the world in 2022. In terms of classification, Wordle and Crossword are two types of crossword games.

Wordle seems to have been greatly simplified in both difficulty and form: the game is updated daily, and players try to solve the puzzle by guessing a five-letter word in six or fewer attempts. The game interface is a 5 x 6 square array. After players enter guesses on the keyboard, the game color-codes the letter squares (green: indicates that the letter is in the answer and is in the right place; Yellow: indicates that this letter is in the answer, but not in this position; Gray: indicates that there is no letter in the answer) indicates the accuracy of the guess. Our model is to model fit and analyze the specificity of the results of this game.

1.2 Question restatement

Based on the above background, this paper needs to establish a reasonable mathematical model to solve the following problems:

Problem 1: Since the number of reported results changes daily, develop and use a model to account for this change and create a forecast interval for the number of reported results on March 1, 2023. Analyze whether any attribute of the word affects the percentage of the player's score in hard mode.

Problem 2: For a given future solution word, develop a model that predicts the relevant percentage of (1, 2, 3, 4, 5, 6, X) at some future date. Explain the uncertainties associated with the model and forecast, and try to explain the forecast of the word EERIE on March 1, 2023.

Problem 3: Develop and summarize a model, categorizing solution words according to difficulty. Identify the attributes of a given word associated with each category. Use the model to analyze the difficulty factor of the word "EERIE" and discuss the accuracy of the classification model.

Problem 4: Illustrate and describe some other interesting features of the data set.

2 Problem analysis

2.1 Question 1

First of all, it is necessary to preprocess the original data provided by the title. Since the wordle heat showed a trend of first surge and then sudden drop in the early stage, the data in the first 120 days or so showed great fluctuations, which was meaningless to the prediction. Only the data after 120 days were analyzed as follows: hampel filtering was used to deal with outliers, the change rule among indicators was studied, and visual charts were drawn to assist.

In order to analyze the statistical law of indicators, we should first use the most basic statistics, such as mean value, variance, etc. Then, the stability test of the data was carried out, the discrete data was processed, and the autocorrelation and partial autocorrelation analysis of the relevant data were carried out, and the trailing characteristics were observed. After passing the stationarity test, white noise test is carried out. If the sequence is white noise, it indicates that the data passes.

2.2 Question 2

In order to develop a reasonable model for a given future solution term, the author first uses the S-W method to detect the normality and variance of the data, and then analyzes the variance. Secondly, the relevant parameters are calculated, and the influencing factors to measure the difficulty of word are mainly divided into the following categories: the frequency of word use after excluding useless data, the number of letters appearing, the letter combination scheme in the word and the guessing strategy of human

2.3 Question 3

According to the pre-processed data, K-Means clustering algorithm is used for it. Since it is an Unsupervised Learning method and does not need pre-labeled training set, it can be obtained by category annotation.

2.4 Question 4

Some other interesting features can be found by observing the original data set. For example, the number of people who have 1 try and 2tries is far less than that of other tries. The author will think from the perspective of information entropy, which will be described in detail later.

3 Problem Hypothesis

1. It is assumed that within the coverage of the data, there is some regularity in the change of the number of daily wordle players, which will not change in the future period of time.

2. Assume that the word is the only thing that affects the number of attempts a person needs to make.

3. Suppose that the proportion of different attempts per day conforms to some parameterized distribution law.

4. It is assumed that the error caused by describing the discrete problem by continuous function is negligible.

5, assume that the vast majority of people are seriously finish the topic, cheating, random answer number can be ignored.

4 Symbol Specification

Symbol	Definition
S_{ob}	Familiarity parameter
S_{re}	Letter repetition factor
$\overrightarrow{S_{vo}}$	Vowel consonant combination
S_s	Confusibility index
S_{abr}	Abnormal shape parameters
f	Word frequency
μ	Sample mean of number of attempts
σ	Sample average of attempts
y_t	Number of people who play wordle on day t

5 Model Establishment And Solution

5.1 Problem 1: Model establishment and solution

In order to predict the number of wordle players in 2023.3.1, we need to find the rule of the number of daily wordle players changing with time since 2022.1.7, which is obviously an independent variable, time and dependent variable, number model. Therefore, we adopt the thought of time series analysis and use ARIMA model to solve and analyze the problem.

5.1.1 Head count data analysis and preprocessing

First, we draw the curve of the number of daily wordle players since 2022.1.7 (blue curve) and the difference curve of the change of the number of daily Wordle players compared to the previous day (orange curve), as shown in Figure 5-1.

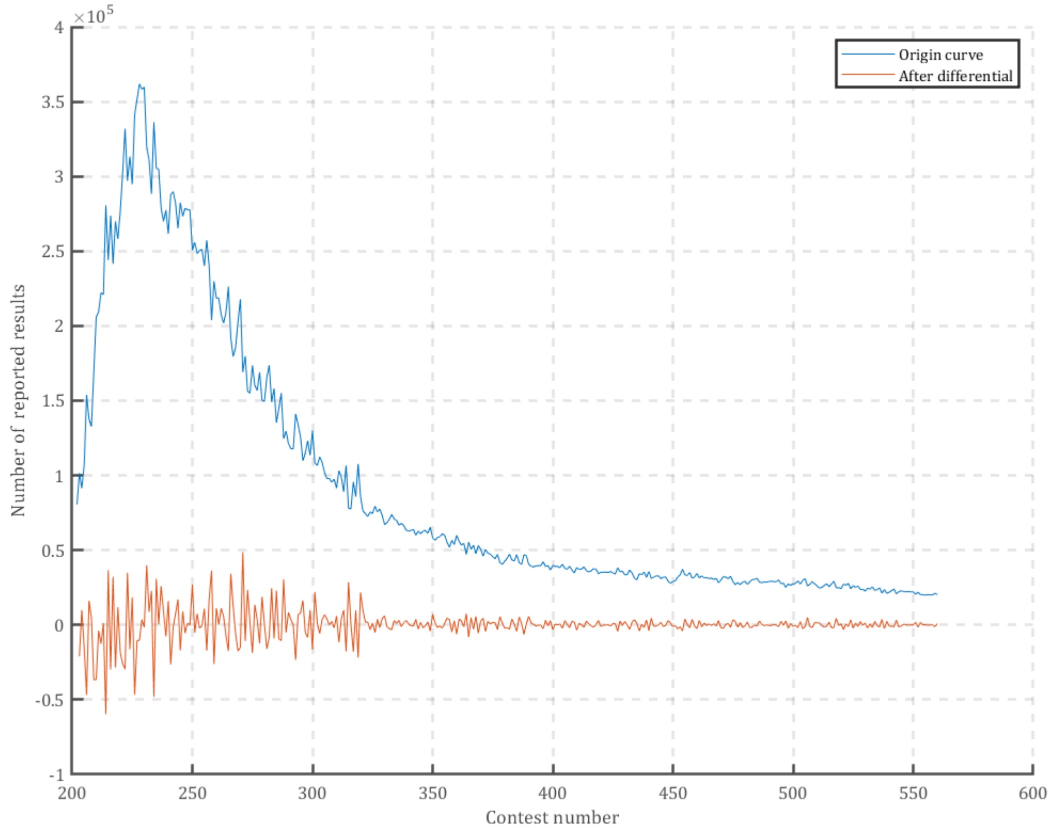

Figure 5-1

As the popularity of wordle shows a surge trend in the early stage, the number of people who play wordle every day shows great fluctuations, and after the popularity, the number of people who play wordle gradually becomes stable. It is obvious that March 1, 2023.1 is after the wordle heat wave, so the first few days with large fluctuations in the number of people are not of analytical significance. In the following model, we only analyze the data after 120 days.

After determining the analysis target, we used Hample filtering to correct some outliers. The basic principle is as follows:

(1) Generate a window of length 5 and calculate the standard deviation σ_i and median of the data m_i in the window.

(2)For an element in the window, if there is $|X_{ij} - m_{ij}| \geqslant 2\sigma_i$, it is replaced with m_i .

Using hampel filtering, we found the outlier at contest number 529 and corrected it to the median of the window it was in.

5.1.2 Stationarity test

Before selecting the ARIMA model, the data should be tested for stationarity first, and the difference should be made for the unstable data until it is stable. Here, ADF tset is used in our stationarity detection method. Its basic principle is to judge whether the sequence is stable by judging whether there is unit root (if the sequence is stable, there is no unit root; Otherwise there will be unit roots).

The raw data ADF test has been passed, proving that its performance is smooth. Therefore, ARIMA analysis can be performed directly without any first-order difference. Figure 5-2 shows the original data (blue curve) and first-order difference data (orange curve).

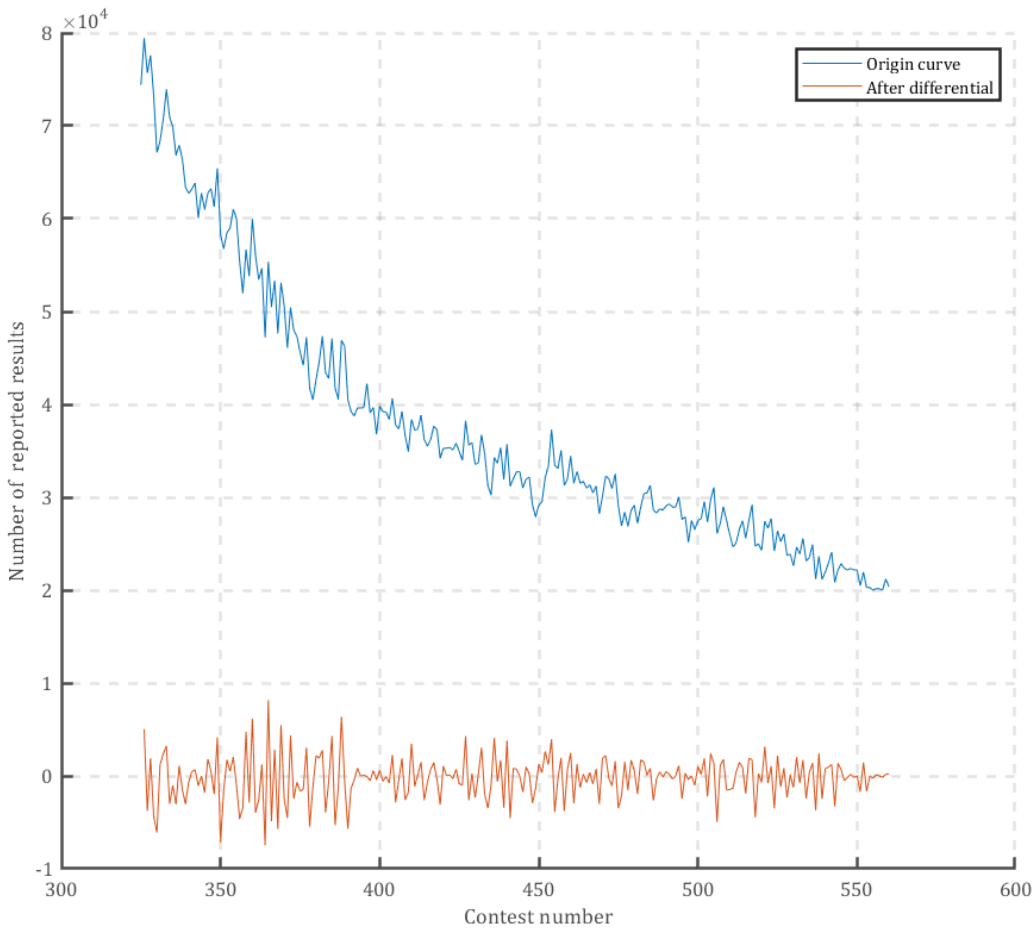

Figure 5-2

5.1.3 Autocorrelation and partial autocorrelation analysis

After the stationarity test passed, we carried out autocorrelation analysis and partial autocorrelation analysis on the data after first-order difference, and the results were shown in Figure 5-3 and Figure 5-4 respectively.

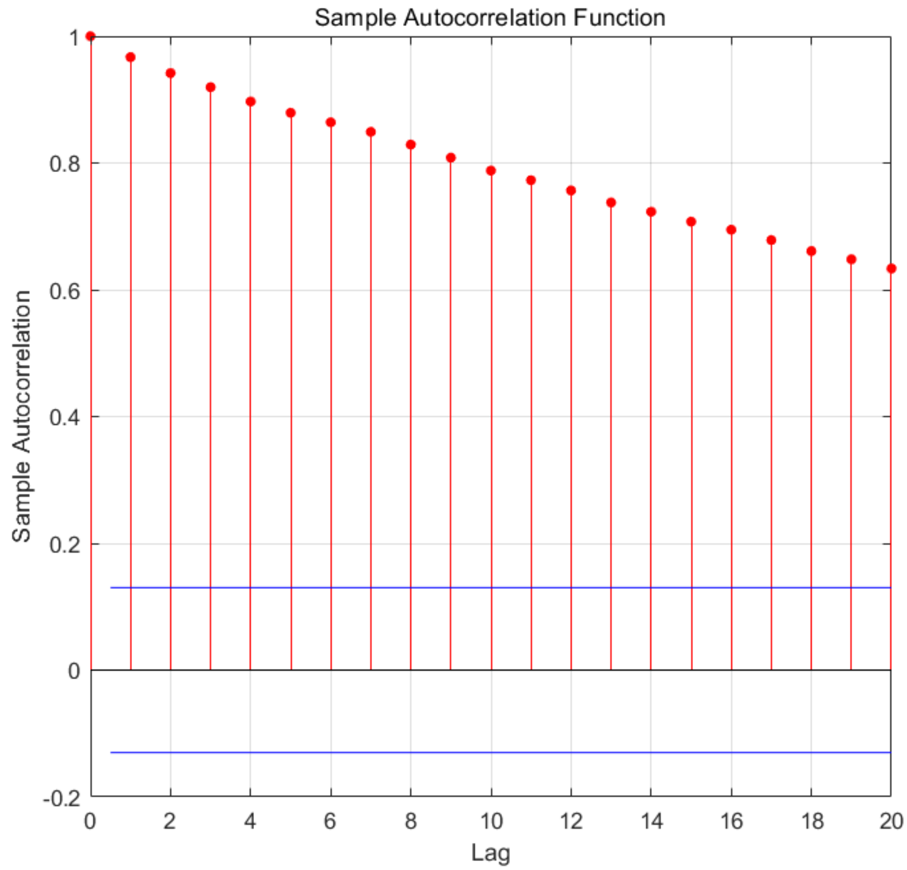
Sample Autocorrelation Function
Sample Autocorrelation
Lag
1
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20

Figure 5-3

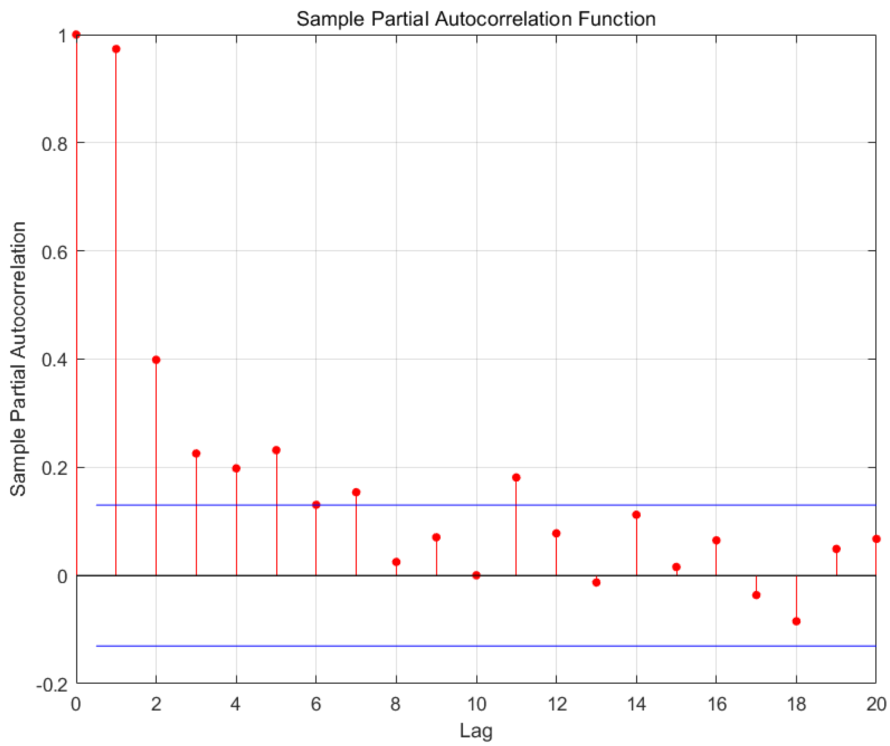

Figure 5-4

5.1.4 The establishment of ARIMA model

In the standard ARIMA(p,d,q) , pis the number of autoregressive terms, qis the number of sliding average terms, and the d is difference number (order) made to make it a stationary sequence. Here, the order has been obtained in the analysis of 5.1.2, and 0 can be taken. As for and, we need to select according to the results of autocorrelation and partial autocorrelation analysis.

It is not difficult to find from FIG. 5-3 and FIG. 5-4 that autocorrelation shows obvious trailing feature, while partial autocorrelation is an insignificant third-order trailing feature. Therefore, the model is chosen, and its mathematical expression is as follows:

$$y_t = u_t + \phi_1 u_{t-1} + \phi_2 u_{t-2} + \ldots + \phi_q u_{t-q} + \theta_1 x_{t\text{-}1} + \theta_2 x_{t\text{-}2} + \ldots + \theta_p x_{t\text{-}p}$$

Where, since q=0, only the latter part of the formula representing autoregression can be retained.

5.1.5 The solution of ARIMA model

matlab built-in function is directly used to solve the sequential model ARIMA(3,0,0), and the known daily wordle players data is brought in. The ARIMA(3,0,0) model is solved as follows:

ARIMA(3,0,0) Mdoel (Gaussian Distribution)				
	Value	Standard Error	T Statistic	P Value
Contest	450.72	408.86	1.1024	0.2703
AR{1}	0.47943	0.05244	9.1425	6.1006E-20
AR{2}	0.27963	0.056755	4.9296	8.3543E-07
AR{3}	0.21909	0.05137	4.2649	1.9997E-05
Variance	4.4151E+06	0.011871	3.7193E+08	0

Table 5-1

The formula it represents is:

$$y_t = 450.72 + x_t + 0.4794x_{t-1} + 0.2796x_{t-2} + 0.2191x_{t-3}$$

According to the model obtained, we can predict the number of people who will play wordle on the day of 2022.3.1 -- 20,459. Meanwhile, in order to make the result more intuitive, we draw the number of people who will play wordle in the next 60 days in Figure 5-6, where the change of the number of people who will play Wordle in the future can be clearly seen.

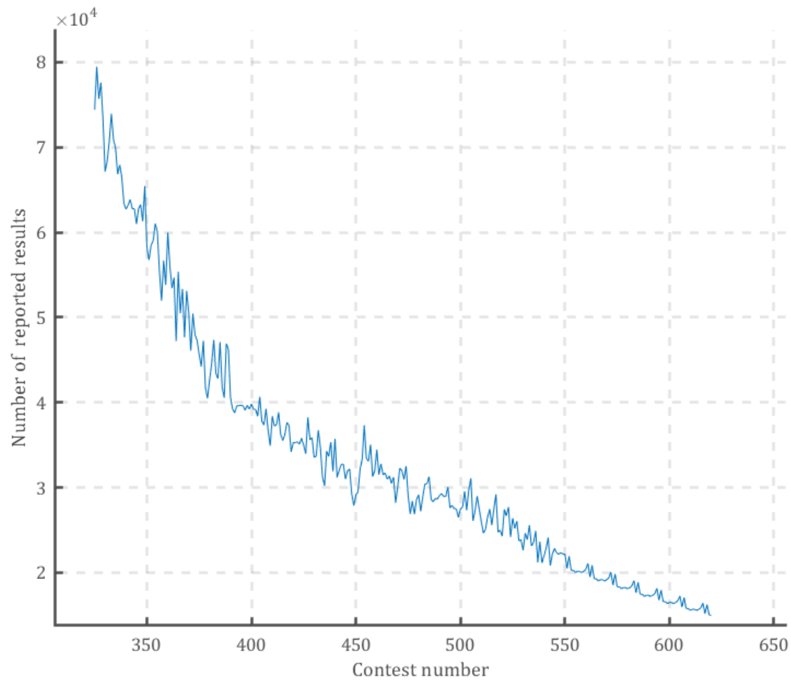

Figure 5-6

For interval, we simply selected the 95% confidence interval range of $ARIMA(3,0,0)$, and the final prediction was 18,829 to 22,289 people.

5.1.6 Residual white noise test

In order to test the reliability of solving the model, we conducted residual white noise test on the model, calculated the residual sequence of the sequential model we obtained, and conducted Ljung-Box white noise test. The results obtained are shown in Table 5-2:

Name	Value
h	1
P	7.11E-07
stat	66.3459
C	31.4104

Table 5-2

The results show that the residual sequence can be considered as white noise sequence at 95% confidence level, which proves that the time series model we solved is effective.

5.1.6 Hard mode ratio

As for the proportion of people who play in difficult mode, we used the parameters in the second question, and used regression analysis to find the correlation coefficient between the proportion of people playing in difficult mode and words. The result is listed below.

Name	Value
R	0.0207
F	0.8101
P	0.6072

Table 5-2-1

It can be clearly found that since the R is pretty low, the proportion of people playing in difficult mode is not affected by words. This is a normal phenomenon, after all, the target word information is unknown before the mode is selected.

5.2 The establishment and solution of Problem Two model

In the second question, in order to find the rule of the number of attempts required by players to face different words, we confirm the distribution of the number of attempts on the one hand, and strive to simplify the solving process; On the other hand, the characteristics of different words are extracted and quantified. Finally, multiple linear regression is used to solve the problem.

5.2.1Determine the parameterized distribution of the number of attempts

Considering that it would make the solving process too tedious to solve the variation rule of each attempt times with words, we will directly try to establish the parameterized distribution of each attempt times here, and express the distribution of each attempt times of all words with uniform distribution expression. Then, in the follow-up analysis process, by abstracting the characteristics of the words, the parameters of the expression of the attempt number distribution of different words were solved to achieve the purpose of solving the target word prediction attempt number distribution.

5.2.1.1 Data preprocessing

In order to solve the parameterized distribution of the number of attempts, it is necessary to convert the frequency into the sample distribution. Taking the word abbey as an example (as shown in Table 5-1), we convert the data of the frequency of attempts into sample data, that is, the data in the fifth to eleventh columns in the table are used as 1,2,2,3,3... (13x3), 4... (29x4)...... This is the sample data representation. The number of attempts for any word

$\vec{T}=(a_1,a_2,a_3,a_4,a_5,a_6,a_7)$, We can turn that into sample data

$$\vec{\tau}=\Big(\underbrace{1,\ldots}_{a_1}\underbrace{2,2,\ldots}_{a_2},\underbrace{3,3,3\ldots}_{a_3}\ldots\Big)$$

2	Word	Number of reported results	Number in hard mode	1try	2tries	3tries	4tries	5tries	6tries	7or more tries(X)
52	abbey	132726	3345	1	2	13	29	31	20	3

Table 5-3

It should be noted that the sample data is not always 100 due to the rounding of the original frequency data. In order to facilitate the subsequent normality test and variance analysis, the following measures were taken to keep the sample data at 100:

(1) When the sample data is less than 100, we supplement it to 100 samples by adding the median of the original sample data.

(2) When there are more than 100 sample data, we delete the median of original sample data to reduce it to 100 samples.

Meanwhile, frequency data of sample data significantly greater than or less than 100 are identified as abnormal data, as shown in the following table. Such abnormal data will not be considered in future analysis.

2	Word	Number of reported results	Number in hard mode	1try	2tries	3tries	4tries	5tries	6tries	7 or more tries(X

279	Nymph	165468	9935	1	2	18	44	26	26	9

Table 5-4

（The sample size is 126）

5.2.1.2 Distribution normality test

It is not difficult to observe the frequency intuitively, and the number of attempts seems to conform to the normal distribution. However, it is obviously inappropriate to draw conclusions based on intuitive feelings. Therefore, Shapiro-Wilk method is used to test whether the number of attempts of each word is in line with the normal distribution. The test results are shown in Table 5-5 and see Attachment 1 for details:

Shapiro-Wilk 检验

F	P Value
0.889350235	0.271291018
0.899630427	0.328748673
0.880615175	0.2291926
0.889306962	0.271067679
0.899950862	0.330683559
0.948080003	0.712193847
0.895090044	0.302279502
0.861708581	0.156809673

注：p 值（显著水平 0.05，p 小于 0.05 拒绝原假设，不正态分布）

Table 5-5

The test shows that the number of attempts of all words, except for a few words, is normally distributed within the confidence range. For a few words whose attempt times do not conform to the normal distribution, the reason is that the maximum attempt times can only be 6, and the attempt times greater than 7 will be truncated to 7, which will result in the left-biased normal distribution of the attempt times of a few difficult words. The word parer shown in the following table:

2	Word	Number of reported results	Number in hard mode	1try	2tries	3tries	4tries	5tries	6tries	7 or more tries(X
279	parer	37309	4130	0	0	4	11	15	22	48

Table 5-6

However, this truncation effect is precisely the reasonableness of treating the number of attempts as a normal distribution, so we treat the number of attempts of all words as a truncated normal distribution.

After confirming the distribution, we will test whether the mean A and standard deviation A of different words are the same.

5.2.1.3 Variance test

To determine whether different words will affect the sample variance of the number of attempts, we conducted levene test on relevant data. Specifically, we conducted variance test for every three words, that is, 1~3 was a group of data, 2~4 was a group of data, 3~5 was a group of data, and so on... Some of the results are shown in Table 5-7 (see Annex 2 for details).

Test statistic	P value
0.00062709	0.999373128
0.002202239	0.997800453
0.01793722	0.982240228
0.16285297	0.85095408
0.103389348	0.902307414
0.069697727	0.932926149
0.014049955	0.986059087
0.04903717	0.952272485

Note: p value (significant level 0.05, p less than 0.05 rejects the original hypothesis and does not have homogeneity of variance)

Table 5-7

According to the test, the variance of all word attempt times has no significant difference within the confidence range. Therefore, when predicting the attempt times of unknown words, we do not need to solve the variance, which can be replaced by the variance mean of the data of known attempt times.

5.2.1.4 Variance analysis

Finally, we need to determine whether the sample mean of the number of attempts is the same for each group, and conduct one-way analysis of variance for the sample data of the number of attempts for all words. The results are shown in Table 5-8.

Name	SS	df	MS	F	P Value
Value	40755.5	99	411.672	1483.71	0
Error	9905.4	35700	0.277		
Total	50660.9	3.58E+04			

Table 5-8

The P value is less than the significant level of 0.05, so it is considered that the sample data mean values of the attempt times of different words are different, which can also be clearly seen in the box graph (only three words are selected) in Figure 5-7. So the mean is going to be the focus of what we're going to do.

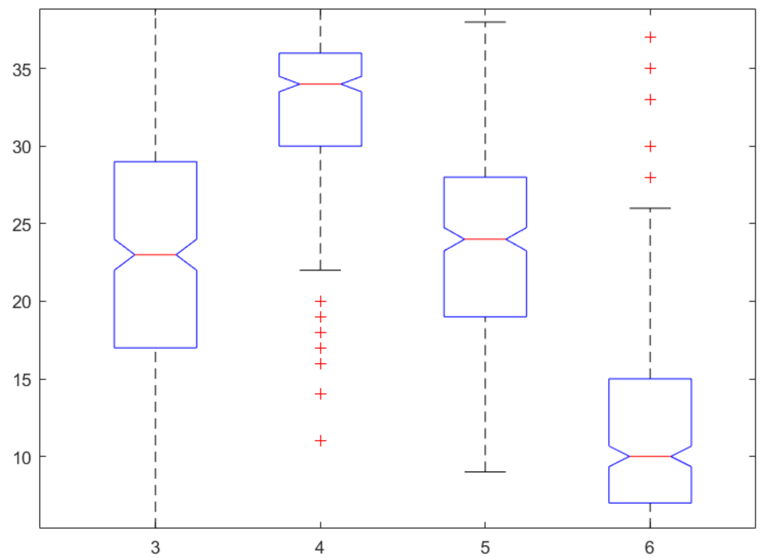

Figure 5-7

5.2.2 Calculation parameter

According to the above test, we have confirmed that the attempts of different words all conform to the normal distribution, and the variances of the attempts of different words are the same. Therefore, as long as the mean of the attempts of target words can be determined, we can predict the proportion of different attempts of target words. It also happens to have a practical meaning -- word guess difficulty. So we're going to use five criteria to measure the difficulty of guessing target wordsc

5.2.2.1 Preprocessing word

Firstly, we found the total word list of all five-letter words on Mathematica and their frequency of use (12,972 words in total). By searching for words with history of wordle on the total list, we found marxh, clen, tash and Rprobe do not appear in the total list. Therefore, it is regarded as a typo and not taken into account in the subsequent solution process, and also treated as naive according to the wordle rule.

5.2.2.2 Familiarity parameter S_{ab}.

It is not hard to see that the more familiar wordle's answer is, the harder it will be for people to guess the word, and the difficulty of guessing will naturally increase. However, directly using word frequency as familiarity score will lead to a huge difference in dimension. For example, the frequency of "their" is an order of magnitude, while that of aahed is an order of magnitude. In the subsequent multiple linear regression, the existence of the latter can almost be ignored, which is obviously not what we want to see.

Therefore, we use the sigmoid function to process the word frequency and normalize all of it to the middle. That is:

$$S_{ob} = \frac{1}{1 + e^{-kf+b}}$$

In this equation, represents the familiar degree parameter of the word, represents the frequency of the word use, and represents the parameter we need to adjust.

In order to have some practical meaning, such as the percentage of people who know the word, we chose the parameters and according to the following adjustment criteria:

(1) 10% of the words (1200 words) scored more than 0.99

(2) 20% (2400 words) scored more than 0.9

(3) 5% (600 words) scored less than 0.01

The function adjusted according to the above criteria is shown in the following formula, which is the familiarity parameter S_{ab} of the word

$$S_{ob} = \frac{1}{1 + e^{-2000000f+4.6}}$$

5.2.2.3 Letter repetition coefficient S_{re}

In guessing words, people often find it difficult to take into account the possibility of a letter repeating itself due to misconceptions, and the difficulty increases exponentially the more letters are repeated. Therefore, we calculate the letter repetition coefficient as follows to measure the degree of letter repetition:

$$S_{re} = \sum_{i=1}^{n} R_i^2$$

In this equation,， S_{re} is the number of times the letter is repeated, and R_i is the number of times the i letter is repeated.

5.2.2.4 Meta-consonant combinations $\overrightarrow{S_{vo}}$

According to morphological knowledge, a word composed of five letters has at most two phonetic knots, among which phonetic knots must contain vowel knots -- fragments of letters consisting of one to three letters (such as eir, oun, etc.) starting with a vowel, separated by consonants.

In the process of processing, in order to coordinate the relationship between vowel knots and vowels, we make the following instructions:

(1) If there is no regular vowel after r, r is regarded as a vowel;

(2) If there is no regular vowel after y, y is regarded as a vowel

(3) If there is no regular vowel after l, l is regarded as a vowel

According to the above law from back to front, you can get a specific five-letter word meta consonant combination vector. It is expressed as:

$$\overrightarrow{S_{vo}} = [\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5]$$

$$\begin{cases} \alpha_i = 0, \ when\ ith\ letter\ is\ consonant \\ \alpha_i = 1, \ when\ ith\ letter\ is\ vowel \end{cases}$$

In the formula, $\overrightarrow{S_{vo}}$ represents the combination vector of the word's consonants, $a_1 - a_5$ Represents whether the first letter in the word is a vowel or a consonant. The calculated partial consonant combinations are shown in the table below.

	3	4	5	6	7
manly	0	1	0	0	1
molar	0	1	0	1	1
havoc	0	1	0	1	0
impel	1	0	0	1	1
condo	0	1	0	0	1
judge	0	1	0	0	1
extra	1	0	0	0	1
poise	0	1	1	0	1
aorta	1	1	1	0	1
excel	1	0	0	1	1
lunar	0	1	0	1	1
third	0	0	1	1	0
slate	0	0	1	0	1
taper	0	1	0	1	1
chord	0	0	1	1	0
manly	0	0	1	0	1
molar	0	1	0	1	1

Table 5-9

5.2.2.5 Confusability index S_s

Through playing wordle in person, we find that if the target word has many close words with similar vowel knots, it is more difficult to guess. This is especially obvious in the hard mode, such as watch/catch, baker/parer, etc.

Thus, we consider the possibility of confusion if a word has a highly repetitive letter segment -- defined here as 3/4 of the letters in the same position -- with other words. At the same time, the more familiar the other word, the more likely people were to confuse it with the other word. According to the above logic, we can calculate the confusability index of each word by following the formula:

$$S_s = \sqrt{k_3} + k_4$$

$$\begin{cases} k_3 = \sum_{i=1}^{n} S_{ob}(i) \\ k_4 = \sum_{j=1}^{m} S_{ob}(j) \end{cases}$$

In the formula, S_s stands for confusion index of words, k_3, k_4 re summed from the familiarity l of three - and four-letter repeated words, respectively

5.2.2.6 Morphological anomaly parameter S_{abr}

morphology We found, by exploring wordle personally, that there were some difficulties in guessing words that violated morphology. Here we measure the degree of morphological anomaly by the rarity of the letter connection.

First of all, according to the total list of 12972 words, we can calculate the weighted frequency of consecutive letters as follows:

$$M(\varepsilon, \delta) = \sum_{i=1}^{12972} S_{ob}(i) \cdot l(i)$$

In the formula, represents the weighted frequency of the letter followed by the letter, represents the frequency of the letter followed by the letter in the first word, is the familiarity of the word. It is worth mentioning that in order to reflect the morphological structure of the beginning and end of the word, we introduce the virtual letter "blank", and fill the virtual word before the first word and after the last word. In the description, if the letter is followed by a virtual letter, it means the letter is at the end of the word; If the letter is followed by a dummy letter, it means the letter is at the beginning of the word.

The calculated data (as shown in Table 5-10 and 5-11) (rows/columns represent, and 27 rows/columns represent virtual letters) :

	1	2	3	4	5	6	7	8	9	10	11
1	0.37	21.45	50.37	43.39	2.03	9.46	22.49	3.07	56.1	1.29	28.35
2	22.24	6.63	0.02	0.01	43.78	0	0	0.18	20.58	0.02	0.02
3	44.22	0	1.76	0	43.49	0	0	77.35	9.46	0	47.69
4	23.82	0.01	0	8.01	69.52	0	10.99	0.23	21.16	0.07	0.76
5	98.28	6.54	11.9	113.18	54.62	8.34	10.63	0.37	14.37	0.09	6.27
6	22.44	0	0	0.01	19.56	8.93	0	0	24.41	0.04	0

7	25.61	0.74	0	0.01	47.13	0	2.52	18.95	15.56	0	0
8	54.07	0	0.01	0.01	50.25	0	0	0	34.32	0	0
9	14.54	5.87	49.22	43.53	32.31	13.55	23.06	0.69	0.19	0.1	6.57

Table 5-10

	12	13	14	15	16	17	18	19	20	21
1	103.95	38.58	99.32	2.48	28.58	0.09	112.36	74.68	59.68	16.04
2	30.15	0.01	0	36.17	0	0	31.09	9.65	2	20.76
3	26.88	0.02	0.03	44.48	0	0	36.58	0.76	10.99	15.31
4	4.32	1.32	0.01	24.16	0	0	25.94	46.53	1.04	8.57
5	64.42	17.92	75.68	3.22	17.2	1.25	135.61	156.4	36.21	3.03
6	27.68	0	0	18.56	0	0	16.63	3.31	17.41	9.34
7	14.34	1.38	3.29	17.9	0	0	36.6	14.73	0	12.76
8	0.09	0.09	1.03	48.26	0.01	0	4.72	5.27	11.48	11.02
9	66.33	19.55	125.3	14.87	12.48	0.06	37.76	43.59	52.16	1

Table 5-11

Then, the target word connection anomaly is calculated according to the following formula:

$$S_{abr}=\sum_{i=1}^{6}\log\left(M\left(\alpha_i,\alpha_{i+1}\right)\right)$$

5.2.3. Solution of model

After determining the parameters to measure the words, we calculated the above parameters for each word of wordle and regarded them as independent variables affecting the mean number of attempts. We expressed them as, and:

$$\vec{S}=\left[S_{ob},S_{re},\vec{S}_{vo},S_s,S_{abr}\right]$$

Multivariate linear regression of dependent variable μ is performed on the independent variable $\vec{S}$ above, and the expression obtained is:

$\mu=4.939+0.040S_{ob}+0.328S_{re}+0.124S_s-0.1088S_{abr}+[0.148,0.195,0.095,0.082,0.139]^T\cdot\vec{S}_{vo}$ The regression results are shown in Table 5-13:

R	F	P	STATS
0.4644	33.2308	0.0000	0.1536

Table 5-13

It can be seen that although the correlation of our regression model is weak, the significance F is very high, which proves that although we cannot find all the parameters to measure the mean value, the parameters found can significantly affect the model, which proves that our model has good generalization and our parameter selection is effective.

Meanwhile, residual analysis of all words in wordle is obtained, as shown in Figure 5-9, where the green line means that the residual analysis passes, and the red line means that the residual analysis fails:

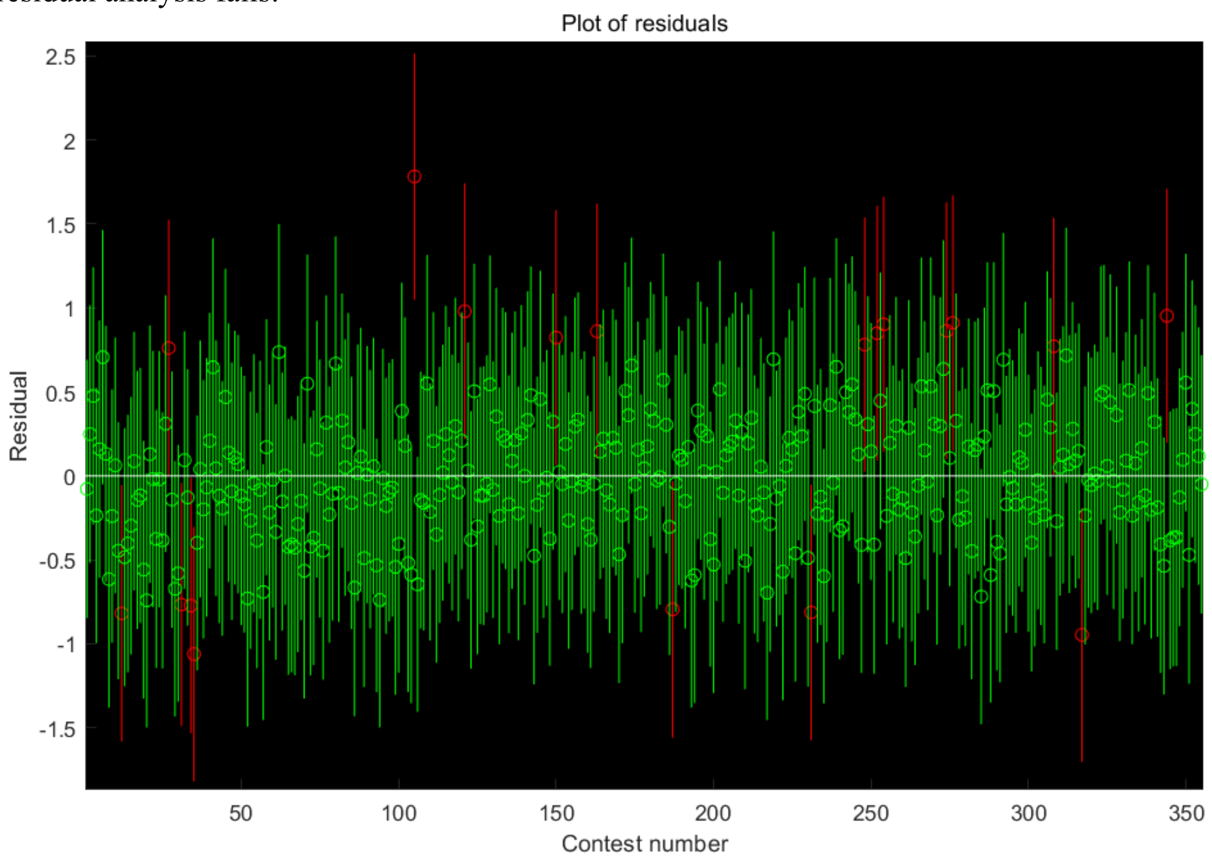

Figure 5-8

It can be seen that our estimation of most words meets the accuracy requirements. In summary, we believe that this multiple regression model has good generalization and high prediction accuracy, which meets the application standards.

5.2.4 Conclusion

It can be concluded from the above model that the score of Eerie is 5.128, and the standard deviation is σ=1.1221 by using the average variance of 358 valid words.

To convert our continuous distribution to the discrete prediction, we integral the probability density function within certain interval, which is listed below.

$$\begin{cases} 1\ try = \dfrac{\displaystyle\int_{0.5}^{1.5} f\,dx}{\displaystyle\int_{0.5}^{+\infty} f\,dx} \\ \vdots \\ 6\ tries = \dfrac{\displaystyle\int_{6.5}^{5.5} f\,dx}{\displaystyle\int_{0.5}^{+\infty} f\,dx} \\ 7\ tries = \dfrac{\displaystyle\int_{7.5}^{6.5} f\,dx}{\displaystyle\int_{0.5}^{+\infty} f\,dx} \end{cases}$$

Among them, $f = \frac{1}{\sqrt{(2\Pi)}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ which means the norm distribution.

The final results are shown in Table 5-14:

Number of tries	Percent
1 try	0
2 tries	1
3 tries	6
4 tries	21
5 tries	34
6 tries	26
7 or more tries	11

Table 5-14

5.3 The establishment and solution of Problem three model

In order to get the label of each type of word, we do K-means cluster analysis on all the word independent variables S obtained before. We divide all words into five categories in the feature space, and the result is shown in Table 5-15 (part), see Attachment 3 for details. Since K-means is an unsupervised algorithm, we need to artificially assign meaning to the labels obtained by clustering. In the table below, category 3 is the simplest word, followed by category 1 and 5. Category 4 is more difficult words, and category 2 is extremely difficult words. According to our prediction, ‘eerie’ could be one of the hardest wordle.

Word	Class
eerie	2
molar	1
havoc	1
impel	3
condo	1
judge	4
extra	1
poise	3
aorta	1
excel	4

Table 5-15

5.4 The establishment and solution of problem four model

Looking at the original data set reveals other interesting features, for example, the percentage of people with 1 try and 2tries is much smaller than the percentage of people with other attempts, why is this?

It is worth thinking about this in terms of information entropy, which is simply the probability of the occurrence of a particular piece of information or the probability of a discrete random event.

$$\mathrm{E[Information]} = \sum_{x} \mathrm{p(x)} * \mathrm{Information(x)}.$$

If a probability distribution is made using a list of common words (there are approximately 2315 words), then the initial uncertainty will be

$I = \log_2(1/p) = \log_2(2315) = 11.17$bits.

and the maximum expected information that can be obtained from the first two steps is approximately

$E[I_1]+E[I_2]=5.77+4.24=10.01$bits.

This means that even with the best strategy, after two guesses, there would be 1.16bits of uncertainty remaining. Given the number of words available, it is unlikely that after just two guesses there will be enough information to ensure that the third time will be the correct answer, so a wild guess is that no algorithm should be able to reduce the average number of guesses to less than 3. This is why the percentage of people with 1 try and 2tries is much smaller than the percentage of people with other numbers of attempts.

6 Evaluation of model

6.1 ARIMA Timing model

(1) Advantages: The model is very simple and only requires endogenous variables without the help of other exogenous variables.

(2) Disadvantages: in essence, it can only capture linear relations, but not nonlinear relations.

6.2 Normality test model based on S-W method

(1) Advantages: more accurate and sensitive than subjective observation.

(2) Disadvantages: Shapiro-Wilk tests are far less useful from a practical point of view than graphical tools for intuitive judgment.

7 Model extension

7.1 ARIMA time series model based on population prediction

(1) ARIMA time series model algorithm can be used to simulate the outbreak trend of some transmissible diseases, predict the number of epidemic cases, and carry out key prevention and control in key areas.

(2) It can be referred to the construction of media drainage, and the model can be promoted and used according to the communication characteristics of media.

7.2 Normality test model based on S-W method

(1) Non-local mean image denoising using Shapiro-Wilk similarity measure.

(2) Mapping disease outbreaks over time can be used to model disease analysis.

8 The letter

Dear Puzzle Editor of the New York Times,

It's our great honour to have the opportunity to write to you. We are a modelling team from China and after an in-depth study of the popular puzzle game, wordle, we have found several useful conclusions as follows.

Firstly, we preprocessed the original data of the wordle game from January 7, 2022 to December 31, 2022, changed the 545th word to probe, corrected the outlier with the Hample filter, found and corrected the outlier at 529. Then we use the ADF stationarity test after the first order difference of the original data, carry out autocorrelation and partial autocorrelation analysis, and finally decide to use the MA (1) model, and the model passes the residual white noise test. After in-depth analysis and inspection, the prediction result is: the number of results reported on March 1, 2023 is 20251, with up and down fluctuations of 1630, that is, the prediction range is (18621,21881).

Secondly, in order to predict the relevant percentage of (1, 2, 3, 4, 5, 6, X) of a future date, we build an evaluation index of the report results based on the normality test and the analysis of variance, and analyze the data. First, we use Shapiro-Wilk method to test and find that the data are consistent with normal distribution, so we can use normal distribution to fit the proportion of different attempts. In order to determine whether different words will affect the variance, we conduct a homogeneous test of variance, and the data all pass. We only need to determine the mean value of the parameter u, and we have analyzed the frequency of use of words, the number of letters, the combination of vowels and consonants, and the degree of connection oddity, and take the word Eerie as an example to predict the percentage of the number of attempts. The results are as follows:

Number of tries	Percent
1 try	0
2 tries	1
3 tries	6
4 tries	21
5 tries	34
6 tries	26
7 or more tries	11

Table 8-1

Thirdly, in order to classify the solution words by difficulty, we use Kmeans clustering analysis. Take the five factors: word use frequency, the number of letters appearing, the combination of vowels and consonants, the combination scheme of letters, and the connection oddity as the influencing parameters. We divide words into five categories and the word Eerie belongs to the second category. The difficulty score is 5.128.

Fourthly, Looking at the original data, we also found some other interesting features, such as: the proportion of people who tried 1 try and 2 tries is far less than the proportion of people who tried other times. Thinking from the perspective of information entropy, which is the probability of the occurrence of certain specific information.

$$\mathrm{E[Information]} = \sum_{x} \mathrm{p(x) * Information(x)}.$$

If we use the common word list for probability distribution (about 2315 words), the initial uncertainty will be

$\mathrm{I} = \log_2(1/p) = \log_2(2315) = 11.17$bits.

and the maximum expected information obtained in the first two steps is about

$E[I_1]+E[I_2]=5.77+4.24=10.01$bits.

This means that even if the best strategy is used, the uncertainty of 1.16 bits will remain after two guesses. Given the number of optional words, it is not possible to get enough information after only two guesses to ensure that the third time is the correct answer, so there should be no algorithm to reduce the average number of guesses to less than 3. This is also why the proportion of 1 try and 2 trials is far less than that of other attempts.

That's all we want to say. Thank you for taking the time to read our letter. We sincerely hope that you can consider our conclusions carefully and we are looking forward to your reply.

9 References

[1]reference.wolfram.com/language/ref/WordFrequencyData.html

[2] What's the Hardest Answer in Wordle? - YouTube

[3]https://www.bilibili.com/video/BV1zZ4y1k7Jw/?spm_id_from=333.999.0.0&vd_source=ff4c03272161f3f1bb53bace1439a711

[4] Blevins, James P. "Word-based morphology." *Journal of linguistics* 42.3 (2006): 531-573.

[5] Ackema, Peter, and Ad Neeleman. *Beyond morphology: Interface conditions on word formation*. Vol. 6. OUP Oxford, 2004.

10 Appendix

A.1：

0.000627090301003	0.999373128111107	0.053950471698113	0.947631645288020
0.002202238942925	0.997800453006087	0.034968431277319	0.965701330854738
0.017937219730942	0.982240228386667	0.024865591397849	0.975474456136979
0.162852969814994	0.850954079934809	0.049476857598865	0.951856152206028
0.103389347764169	0.902307413524010	0.523990860624524	0.600913353701457
0.069697726704971	0.932926148920000	0.048178807947020	0.953085835260610
0.014049955396967	0.986059087148218	0.015934539190353	0.984205610760647
0.049037169726825	0.952272484745058	0.010685210312076	0.989377944387556
0.066666666666667	0.935736868262292	0.012600000000000	0.987487749134692
0.010087862024081	0.989968442488022	0.011791477037650	0.988285397191862
0.187119332280622	0.830937516150105	0.010431918008785	0.989628284217044
0.242521008403361	0.787170059921090	0.010196779964222	0.989860744397317
0.274008350730688	0.763440581108373	0.096435600262410	0.908534329737879
0.179295624332977	0.837333320999928	0.109833585476550	0.896579063261802
0.090799578376751	0.913616294620580	0.204801200300075	0.816681777100779
0.058072750478622	0.943757346069786	0.040429564118762	0.960463760952118
0.018603980386501	0.981586853488871	0.005314960629921	0.994700699228982
0.000396510705789	0.999603576624958	0.102005231037489	0.903543067447606
0.058849557522124	0.943029241090173	0.052881925013990	0.948638794375248
0.068245125348189	0.934271982175265	0.075936377629554	0.927170488212025
0.074279939363315	0.928694829869475	0.024081506637851	0.976237534885267
0.064460232195460	0.937788825130082	0.069690265486726	0.932933056222757
0.046606704824203	0.954577501793658	0.075152335815843	0.927891657953519
0.023115024766098	0.977179037019048	0.063652437484213	0.938541314128449
0.036783777410988	0.963956786785020	0.043970172280792	0.957084962850068
0.060894386298763	0.941115595369780	0.088378566457898	0.915808996884037
0.069206008583691	0.933381483790544	0.043052391799544	0.957959529847644
0.068240123140072	0.934276620440052	0.583264971287940	0.568278359908431
0.029117774880487	0.971347716124114	0.390031369815266	0.682620295834957
0.029224652087475	0.971244242107540	0.264023210831721	0.770878395608401
0.025164361822716	0.975183864314594	0.051928332827209	0.949538598501420
0.037685337726524	0.963091690678269	0.215564202334630	0.808137431983843
0.009518054086720	0.990532081007120	0.224018475750577	0.801495535015489
0.005605058924978	0.994412355007264	0.046719428133776	0.954470459681947
0.000420816383785	0.999579281980701	0.037411526794742	0.963354335898961
0.453083109919571	0.642719238101614	0.030212302667392	0.970288619671548

0.385600386566803	0.685526191475476	0.033419689119171	0.967192442358767
0.282105263157894	0.757467811262974	0.021303575957393	0.978946385789695
0.012302839116719	0.987780829935733	0.010123734533183	0.989932970302828
0.101890756302520	0.903645347571211	0.002401280683031	0.997601919601641
0.166462293071735	0.847943232645948	0.027284263959391	0.973124753137453
0.258712587125871	0.774866990363944	0.026117647058824	0.974257317367828
0.060542280549709	0.941444804532791	0.028009084027252	0.972421826851672
0.417342211928199	0.665008833461942	0.368405365126676	0.696933784458302
0.496180290297937	0.616938712532771	0.244404655326768	0.785727686585886
0.321462639109697	0.729165338303874	0.257455481656297	0.775814504421728
0.076549210206561	0.926607232120304	0.026207880192548	0.974169665832335
0.116770462633451	0.890457918974877	0.671854391981007	0.523112186892541
0.144911979390296	0.866097564952432	0.642208774583963	0.537766579904752
0.014201855709146	0.985909550527132	0.433901192504259	0.654576900653611
0.014373323112304	0.985740781568747	0.020786199095023	0.979451818873075
0.015769554247267	0.984367717714340	0.026336054212229	0.974045173974005
0.002055146429183	0.997947198066871	0.000746825989544	0.999253483776788
0.109507640067911	0.896867827378284	0.000829187396352	0.999171194447833
0.120368004088934	0.887301721175229	0.448937426210153	0.645261611245828
0.270768728578423	0.765844963186478	0.577887537993920	0.571156317792433
0.212087738677607	0.810886357803074	0.652783184118333	0.532487751445621
0.065892796176169	0.936455989721190	0.006903353057199	0.993123048364486
0.037583662261884	0.963189210640463	0.036155747836836	0.964559925494128
0.008033664881408	0.992002073506789	0.017674418604651	0.982497886323174
0.025609756097561	0.974750840910264	0.017790262172285	0.982384300635382
0.055736448956139	0.945950939755284	0.107213114754098	0.898903537894402
0.004315659679408	0.995694669330922	0.069726390114740	0.932899613989648
0.004583151462244	0.995428496379925	0.113303844845185	0.893511060326172
0.009429280397022	0.990619925625950	0.091385522436947	0.913086484233821
0.011606597434331	0.988467890651233	0.747247706422018	0.487803103139284
0.003532974427995	0.996473949996164	1.299068422743330	0.297167351818026
0.000659920809503	0.999340321067394	1.449888641425380	0.260721966409486
0.006598984771574	0.993425142885699	0.023785322219963	0.976525956950595
0.060310880829016	0.941661226565171	0.017503646593040	0.982665356469734
0.073368463721119	0.929534805280415	0.016460069091648	0.983689445867120
0.086052511762028	0.917921205723218	0.016006739679865	0.984134679063846
0.007305819020764	0.992723745706968	0.009914624070504	0.990139767050353
0.015498154981550	0.984634447249080	0.014523449319213	0.985593044307994
0.006871035940803	0.993155119213832	0.009224461906389	0.990822633592259
0.952142857142857	0.404513641554606	0.015789473684211	0.984348144176486
0.909527498063516	0.420441968493282	0.077293318233296	0.925923831428018
1.082012532252110	0.359962689852273	0.984844384303112	0.392745193428987
0.302423797282409	0.742706962276217	1.077592732778190	0.361386026914369

0.258976881455976	0.774667948271059	1.287096774193540	0.300294346677797
0.271601753933453	0.765225906033958	0.033463973470003	0.967149770373002
0.030267480056312	0.970235262430771	0.007612603087333	0.992419492716873
0.002711660138596	0.997292420408461	0.808249232119350	0.461167983144364
0.155938349954669	0.856755393511742	0.669351817159985	0.524331953395183
0.176470588235294	0.839656186502869	0.784894259818731	0.471169699401225
0.157585260682085	0.855369671719684	0.040087754287994	0.960790650534250
0.021179164281626	0.979067898919225	0.039211484893936	0.961629234940825
0.024821109123435	0.975517728972717	0.018818292189397	0.981376946733463
0.056766345666498	0.944983254568774	0.081712062256809	0.921877105489673
0.053071253071253	0.948460258251719	0.090291262135922	0.914076190666457
0.011992893100385	0.988086623307739	0.071223584659536	0.931514762877644
0.004496788008565	0.995514425387152	0.015590200445434	0.984543976533165
0.006916426512968	0.993110074859654	0.044128285507596	0.956934383949993
0.008601676224085	0.991439285048121	0.050694444444445	0.950704296220218
0.001551991722711	0.998449345586236	0.007214206437292	0.992814622549208
0.031962616822430	0.968597631020731	0.003796781775447	0.996211214492069
0.161332462442848	0.852226020611107	0.004607283896446	0.995404486815736
0.203133743704532	0.818014466509167	0.034914950760967	0.965752778694656
0.209141540234300	0.813224124514841	0.329500396510706	0.723530923581437
0.044748577579579	0.956343903848509	0.310309629125553	0.737064451223543
0.045540796963947	0.955590348092444	0.279497395035243	0.759385844450857
0.538831064851881	0.592551590884024	0.093291102792974	0.911365818219522
0.494489346069066	0.617928294128981	0.029712460063898	0.970772122434169
0.501109877913429	0.614063833242661	0.049161585365854	0.952154657291347
0.001172638436482	0.998828125132532	0.313692946058091	0.734658217739957
0.009528130671506	0.990522110426940	0.406028368794326	0.672242594473559
0.007815407517678	0.992218418380220	0.470729366602687	0.632021346118517
0.012884455527847	0.987207289752006	0.004531722054381	0.995479666082457
0.000356083086053	0.999643987345486	0.018853695324284	0.981342276088556
0.001818457342022	0.998183378401725	0.014740290126345	0.985379697500145
0.047111609646663	0.954098147825104	0.035599505562423	0.965094472237358
0.047173343317110	0.954039556557182	0.014554275318375	0.985562711885034
0.194637299497246	0.824842754986778	0.024557956777996	0.975773770045838
0.184364820846905	0.833183090218106	0.296032553407934	0.747315279666790
0.587750716332378	0.565889950093249	0.395705521472392	0.678919080417179
0.622528659245721	0.547746568780102	0.330892448512585	0.722560027540152
0.562359297613687	0.579558211730574	0.047270062293881	0.953947768863148
0.080982094411286	0.922544259491377	0.065762907412749	0.936576749412957
0.002512036843207	0.997491465310334	0.140404040404040	0.869949499146359
0.047919542496549	0.953331656961988	0.009821753364860	0.990231625574927
0.129154583983939	0.879645196576130	0.054789414834141	0.946841739515776
0.127431149836626	0.881141171128983	0.210251046025104	0.812342869484737

0.099409712905822	0.905865257576624	0.431519138755981	0.656066302456082
0.074851485148515	0.928168548943702	0.370675453047776	0.695415738208128
0.069247952345495	0.933342633903606	0.116407982261640	0.890776621662040
0.029184335245697	0.971283273739338	0.050673508659397	0.950724088705761
0.014348574708246	0.985765138515873	0.069213594641528	0.933374457165865
0.030495552731893	0.970014747822610	0.007768666378938	0.992264756762986
0.040480591497227	0.960414971557843	0.005143277002204	0.994871388532645
0.031049778215870	0.969479120780083	0.010877862595420	0.989187584845587
0.334570056743780	0.720002011481455	0.025080385852090	0.975265531790603
0.344720496894409	0.712993778975997	0.023172265133824	0.977123247990726
0.456902848794740	0.640386602031530	0.000598205383848	0.999401993373087
0.073410404624277	0.929496135887073	0.002190980463758	0.997811684043789
0.016871566832331	0.983285491027908	0.004774897680764	0.995237744188985
0.006293706293706	0.993728243364962	0.008734602463606	0.991307632300752
0.063529411764706	0.938655975742147	0.001855287569573	0.998146623259228
0.096805578076861	0.908201823368275	0.019086460032626	0.981114361162150
0.090537780803268	0.913853122276420	0.085714285714286	0.918228787468458
0.008224811514736	0.991812644856256	0.114919354838709	0.892086793414808
0.006805095096842	0.993220561136406	0.111111111111110	0.895448263225421
0.040660972892685	0.960242523491276	0.013338139870224	0.986760162077391
0.065612185120094	0.936716900063712	0.130838079739625	0.878186613542864
0.094302554027505	0.910453980408268	0.183506686478454	0.833884048285025
0.084680173002504	0.919169916995057	0.049163879598662	0.952152484697237
0.064915966386555	0.937364588603473	0.001027045532352	0.998973540235043
0.087251257380276	0.916831991612364	0.015829145728643	0.984309162321558
0.085679063077871	0.918260825410843	0.009409304756926	0.990639693401774
0.109690444145356	0.896705863430511	0.018360655737705	0.981825237223751
0.055598012991976	0.946081097337810	0.036119472100023	0.964594776312100
0.005358615004122	0.994657302872684	0.004154302670623	0.995855269030284
0.010549229738781	0.989512331169538	0.096976744186047	0.908048038754185
0.006374632232756	0.993647884781394	0.066488066244520	0.935902778756969
0.095923709798055	0.908994598866564	0.065535524920467	0.936788192267023
0.098351442487823	0.906813986471747	0.094271056006841	0.910482361029290
0.077704918032787	0.925546052178885	0.092761265580058	0.911843879028327
0.005451713395639	0.994564761763221	0.053030303030303	0.948498870925129
0.007038123167155	0.992989317651683	0.058046683046683	0.943781790028838
0.019657258064516	0.980555704938363	0.143778207607994	0.867064560993138
0.127070783132530	0.881454333687792	0.433890987587695	0.654583273356466
0.096900234436051	0.908116775551950	0.338264299802761	0.717442551273568
0.122280110327919	0.885629251135179	0.306494900697799	0.739788015460459
0.054582763337893	0.947036244020084	0.040360733293909	0.960529577323198
0.058017727639001	0.943808942925942	0.587947882736156	0.565785226142430
0.130155558726987	0.878777626547001	0.514976244577566	0.606056634493092

0.109334032511798	0.897021673324858	0.333802816901408	0.720534845009885

A.2：

0.889350235462188	0.271291017532348	0.895685434341430	0.305650979280471
0.899630427360534	0.328748673200607	0.887027025222778	0.259511351585388
0.880615174770355	0.229192599654197	0.887786269187927	0.263313531875610
0.889306962490081	0.271067678928375	0.893938541412353	0.295843899250030
0.899950861930847	0.330683559179306	0.906352877616882	0.371198326349258
0.948080003261566	0.712193846702575	0.927015721797943	0.525814235210418
0.895090043544769	0.302279502153396	0.924600720405578	0.505971193313598
0.861708581447601	0.156809672713279	0.854901790618896	0.136198922991752
0.935543537139892	0.598964869976043	0.868251204490661	0.179188981652259
0.831759989261627	0.083147980272770	0.931207060813903	0.561211645603179
0.840082287788391	0.099533438682556	0.804889202117919	0.045774161815643
0.807829260826110	0.048917800188065	0.883663594722747	0.243213102221488
0.910477995872497	0.399193584918975	0.909890472888946	0.395115762948989
0.920486271381378	0.473179429769516	0.927403271198272	0.529037475585937
0.785893619060516	0.029623957350850	0.887123465538024	0.259991914033889
0.910773336887359	0.401254534721374	0.859418034553527	0.149582371115684
0.900373935699462	0.333251297473907	0.907319903373718	0.377627819776535
0.898056030273437	0.319371670484542	0.865465164184570	0.169336065649986
0.860861301422119	0.154100731015205	0.887164950370788	0.260198622941970
0.882643520832061	0.238443061709404	0.895371556282043	0.303869694471359
0.894931972026824	0.301389575004577	0.897973477840423	0.318885773420333
0.874633252620697	0.203671708703041	0.881307840347290	0.232316672801971
0.822879552841186	0.068440623581409	0.867182612419128	0.175351724028587
0.916782975196838	0.444821447134017	0.923971533775329	0.500872075557708
0.890266716480255	0.276057541370391	0.863801717758178	0.163683995604515
0.936527609825134	0.607670009136199	0.920680761337280	0.474699288606643
0.901766002178192	0.341809481382370	0.839522421360015	0.098344311118126
0.895371377468109	0.303869009017944	0.900502681732177	0.334035813808441
0.816915392875671	0.059964098036289	0.897116124629974	0.313874602317810
0.890579223632812	0.277698397636413	0.877189993858337	0.214264363050460
0.821572244167327	0.066492192447186	0.913477182388305	0.420473754405975
0.898490250110626	0.321936637163162	0.838123083114624	0.095428794622421
0.902962505817413	0.349299252033233	0.891429662704467	0.282204180955886
0.823694229125976	0.069681569933891	0.916053414344787	0.439368963241577
0.879617393016815	0.224755063652992	0.888294696807861	0.265885740518569
0.882576406002044	0.238131821155548	0.846441745758056	0.113996773958206
0.891493737697601	0.282546103000640	0.871402680873870	0.190937533974647
0.944346725940704	0.678127944469451	0.894160211086273	0.297074228525161
0.886620938777923	0.257496684789657	0.895687520503997	0.305662959814071
0.912549018859863	0.413805723190307	0.860532343387603	0.153060436248779
0.932092607021331	0.568834245204925	0.895868539810180	0.306693613529205

0.892730712890625	0.289212346076965	0.834148287773132	0.087575480341911
0.912540972232818	0.413747876882553	0.892675340175628	0.288911193609237
0.817625164985656	0.060918867588043	0.922136187553405	0.486170798540115
0.853196382522583	0.131431907415390	0.873580157756805	0.199441954493522
0.902277410030364	0.344995409250259	0.920764327049255	0.475353717803955
0.936537623405456	0.607758581638336	0.937852501869201	0.619459807872772
0.890089869499206	0.275132447481155	0.867613494396209	0.176890239119529
0.948811054229736	0.718864142894744	0.908707320690155	0.386995106935501
0.896693944931030	0.311429768800735	0.889678835868835	0.272992342710495
0.890591204166412	0.277761608362197	0.910311222076416	0.398032903671264
0.871231794357299	0.190283328294754	0.877747058868408	0.216633960604667
0.864448189735412	0.165860325098037	0.897682666778564	0.317178815603256
0.885192275047302	0.250511199235916	0.889593601226806	0.272550135850906
0.884843170642852	0.248828560113906	0.868514895439147	0.180146992206573
0.899706065654754	0.329204708337783	0.886620938777923	0.257496684789657
0.885906100273132	0.253981620073318	0.910548865795135	0.399687260389328
0.819781482219696	0.063907064497471	0.875938892364501	0.209023296833038
0.868186771869659	0.178955465555191	0.928207099437713	0.535756111145019
0.873417496681213	0.198795735836029	0.928351044654846	0.536964118480682
0.848863065242767	0.119989521801471	0.881893634796142	0.234986931085586
0.868496060371398	0.180078372359275	0.821297407150268	0.066089190542698
0.895850777626037	0.306592285633087	0.879919171333312	0.226089328527450
0.855581998825073	0.138143375515937	0.916878759860992	0.445540875196456
0.891412973403930	0.282115399837493	0.857035577297210	0.142382204532623
0.919863879680633	0.468335211277008	0.882087945938110	0.235878363251686
0.888944327831268	0.269201636314392	0.934595167636871	0.590621054172515
0.865464925765991	0.169335082173347	0.919055044651031	0.462086647748947
0.900373935699462	0.333251297473907	0.897643268108367	0.316948175430297
0.919594466686248	0.466247916221618	0.918401956558227	0.457080096006393
0.865078270435333	0.168006315827369	0.878853023052215	0.221405282616615
0.892910718917846	0.290192812681198	0.860887587070465	0.154184281826019
0.861049115657806	0.154697462916374	0.897269189357757	0.314764499664306
0.884117245674133	0.245359912514686	0.894629240036010	0.299691051244735
0.889070093631744	0.269847631454467	0.889892399311065	0.274102658033370
0.863510310649871	0.162711024284362	0.906729996204376	0.373695701360702
0.881151854991912	0.231610178947448	0.894511461257934	0.299032002687454
0.912793278694152	0.415553301572799	0.899713814258575	0.329251646995544
0.848983526229858	0.120295062661170	0.909379422664642	0.391593068838119
0.915214061737060	0.433150589466094	0.923503696918487	0.497100323438644
0.932409107685089	0.571569979190826	0.842990159988403	0.105924770236015
0.888177871704101	0.265292793512344	0.903789162635803	0.354546517133712
0.894389510154724	0.298351198434829	0.852950692176818	0.130757853388786
0.868056118488311	0.178482934832572	0.881576836109161	0.233539640903472

0.865755438804626	0.170339658856391	0.880685269832611	0.229507103562355
0.915015399456024	0.431687951087951	0.844962120056152	0.110470205545425
0.832561254501342	0.084609821438789	0.875903367996215	0.208876013755798
0.894796490669250	0.300628155469894	0.915380060672760	0.434376060962677
0.906526386737823	0.372345954179763	0.885506868362426	0.252035856246948
0.893974959850311	0.296045780181884	0.888637304306030	0.267630338668823
0.875275731086730	0.206290096044540	0.888432562351226	0.266586452722549
0.879575848579406	0.224571675062179	0.879851043224334	0.225787550210952
0.878883004188537	0.221535950899124	0.922329783439636	0.487709581851959
0.907062828540802	0.375910758972167	0.934740126132965	0.591893255710601
0.892206907272338	0.286374151706695	0.915866851806640	0.437981218099594
0.892705917358398	0.289077728986740	0.904521107673645	0.359242171049118
0.879737138748168	0.225283652544021	0.905673801898956	0.366731643676757
0.924901127815246	0.508416175842285	0.884766578674316	0.248460561037063
0.910829126834869	0.401644468307495	0.900670111179351	0.335057884454727
0.866935908794403	0.174476176500320	0.908612310886383	0.386348485946655
0.890025258064270	0.274794936180114	0.914134263992309	0.425239413976669
0.879407584667205	0.223831430077552	0.899899840354919	0.330374777317047
0.931595921516418	0.564553141593933	0.872180342674255	0.193937942385673
0.937688112258911	0.617992341518402	0.903843402862548	0.354892760515213
0.824067413806915	0.070257008075714	0.907442092895507	0.378446042537689
0.806718349456787	0.047706846147776	0.896684169769287	0.311373621225357
0.844497025012969	0.109382346272468	0.897287130355834	0.314869284629821
0.870223999023437	0.186466753482818	0.887900888919830	0.263891756534576
0.855340719223022	0.137450844049453	0.926197290420532	0.519042313098907
0.888572633266448	0.267300218343734	0.900483787059783	0.333920478820800
0.875180006027221	0.205897971987724	0.873647093772888	0.199708566069602
0.867588877677917	0.176801964640617	0.914488554000854	0.427824229001998
0.924269437789916	0.503282904624938	0.888572633266448	0.267300218343734
0.881476998329162	0.233084976673126	0.938310623168945	0.623553633689880
0.956870317459106	0.791478037834167	0.903650522232055	0.353662401437759
0.903376638889312	0.351920455694198	0.840049803256988	0.099464125931263
0.958439707756042	0.805241584777832	0.884557723999023	0.247459724545478
0.837500810623168	0.094158045947552	0.908130884170532	0.383082985877990
0.895536422729492	0.304804503917694	0.875993371009826	0.209249183535575
0.869032382965087	0.182040408253669	0.902798652648925	0.348266154527664
0.863510310649871	0.162711024284362	0.933446526527404	0.580577969551086
0.878092944622039	0.218116745352745	0.876726329326629	0.212308943271636
0.933274507522583	0.579080164432525	0.855615854263305	0.138240784406662
0.867700576782226	0.177202567458152	0.860333323478698	0.152433961629867
0.854798078536987	0.135904714465141	0.868336200714111	0.179497241973876
0.911644399166107	0.407377421855926	0.896463274955749	0.310100555419921
0.896480083465576	0.310197383165359	0.905260503292083	0.364033102989196

0.887223422527313	0.260490506887435	0.877381563186645	0.215076640248298
0.881173014640808	0.231705829501152	0.897452712059020	0.315834522247314
0.880804419517517	0.230042353272438	0.920258820056915	0.471405386924743
0.859234988689422	0.149017870426177	0.890560269355773	0.277598828077316
0.849793553352355	0.122367627918720	0.876054883003234	0.209504559636116
0.930122971534729	0.551945686340332	0.929869592189788	0.549790740013122
0.903389632701873	0.352003097534179	0.878259062767028	0.218831643462181
0.898377537727355	0.321269214153289	0.880000710487365	0.226451188325881
0.865785598754882	0.170444354414939	0.867093563079833	0.175035312771797
0.865238904953002	0.168557167053222	0.898278713226318	0.320685297250747
0.886416137218475	0.256485313177108	0.903105258941650	0.350201219320297
0.913272976875305	0.419000506401062	0.885018229484558	0.249670982360839
0.908057212829589	0.382585078477859	0.917628884315490	0.451199650764465
0.925954580307006	0.517043232917785	0.931045353412628	0.559824943542480
0.832295477390289	0.084122434258461	0.873647093772888	0.199708566069602
0.921801328659057	0.483516931533813	0.839784979820251	0.098900251090527
0.870139360427856	0.186149150133132	0.898446738719940	0.321678847074508
0.892675697803497	0.288913160562515	0.875920653343200	0.208947598934173
0.832295477390289	0.084122434258461	0.922108113765716	0.485947906970977
0.859750211238861	0.150611594319343	0.897194802761077	0.314331769943237
0.884045898914337	0.245021253824234	0.882223069667816	0.236499845981597
0.901530265808105	0.340348243713378	0.910438418388366	0.398917913436889
0.868792176246643	0.181159406900405	0.908482909202575	0.385468631982803
0.916572332382202	0.443242460489273	0.878572285175323	0.220185711979866
0.907279968261718	0.377360701560974	0.920726180076599	0.475055038928985
0.887131035327911	0.260029584169387	0.912517189979553	0.413578122854232
0.888560295104980	0.267237156629562	0.907075047492980	0.375992089509964
0.893534541130065	0.293612092733383	0.913338303565979	0.419471204280853
0.900096714496612	0.331566870212554	0.913338303565979	0.419471204280853
0.906906962394714	0.374872207641601	0.873826265335083	0.200423687696456
0.917716383934020	0.451862514019012	0.902399837970733	0.345761626958847
0.908649802207946	0.386603444814682	0.885670065879821	0.252829521894454
0.885006546974182	0.249614804983139	0.898067235946655	0.319437623023986
0.946767449378967	0.700212001800537	0.883282184600830	0.241420015692710

A.3：

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0.3700	21.4500	50.3700	43.3900	2.0300	9.4600	22.4900	3.0700	56.1000	1.2900	28.3500	########	38.5800	99.3200	2.4800	28.5800	0.0900	112.3600	74.6800	59.6800	16.0400
2	22.2400	6.6300	0.0200	0.0100	43.7800	0.0000	0.0000	0.1800	20.5800	0.0200	0.0200	30.1500	0.0100	0.0000	36.1700	0.0000	0.0000	31.0900	9.6500	2.0000	20.7600
3	44.2200	0.0000	1.7600	0.0000	43.4900	0.0000	0.0000	77.3500	9.4600	0.0000	47.6900	26.8800	0.0200	0.0300	44.4800	0.0000	0.0000	36.5800	0.7600	10.9900	15.3100
4	23.8200	0.0100	0.0000	8.0100	69.5200	0.0000	10.9900	0.2300	21.1600	0.0700	0.7600	4.3200	1.3200	0.0100	24.1600	0.0000	0.0000	25.9400	46.5300	1.0400	8.5700
5	98.2800	6.5400	11.9000	113.1800	54.6200	8.3400	10.6300	0.3700	14.3700	0.0900	6.2700	64.4200	17.9200	75.6800	3.2200	17.2000	1.2500	135.6100	156.4000	36.2100	3.0300
6	22.4400	0.0000	0.0000	0.0100	19.5600	8.9300	0.0000	0.0000	24.4100	0.0400	0.0000	27.6800	0.0000	0.0000	18.5600	0.0000	0.0000	16.6300	3.3100	17.4100	9.3400
7	25.6100	0.7400	0.0000	0.0100	47.1300	0.0000	2.5200	18.9500	15.5600	0.0000	0.0000	14.3400	1.3800	3.2900	17.9000	0.0000	0.0000	36.6000	14.7300	0.0000	12.7600
8	54.0700	0.0000	0.0100	0.0100	50.2500	0.0000	0.0000	0.0000	34.3200	0.0000	0.0000	0.0900	0.0900	1.0300	48.2600	0.0100	0.0000	4.7200	5.2700	11.4800	11.0200
9	14.5400	5.8700	49.2200	43.5300	32.3100	13.5500	23.0600	0.6900	0.1900	0.1000	6.5700	66.3300	19.5500	125.3000	14.8700	12.4800	0.0600	37.7600	43.5900	52.1600	1.0000

10	3.1700	0.0000	0.0000	0.0100	7.6200	0.0000	0.0000	0.0100	2.1700	0.0100	0.0000	0.0000	0.0000	0.0100	9.4400	0.0000	0.0000	0.0500	0.0000	0.0000	4.8200
11	3.6500	0.0100	0.0000	0.0000	41.2700	0.0200	0.0000	0.3400	11.4800	0.0000	0.1400	1.1500	0.0000	8.1000	1.0400	0.0000	0.0000	0.4800	41.3500	0.0600	1.5700
12	74.4500	2.6500	1.4800	14.2600	93.6200	1.2700	1.5500	0.0000	53.8900	0.0200	7.0800	63.6300	4.5100	0.2800	72.3400	4.4100	0.0000	0.0000	52.1300	19.7700	24.1200
13	52.9700	8.2100	0.0100	0.0100	50.8200	0.0900	0.0000	0.0200	29.7800	0.0000	0.0200	0.8900	11.9100	0.9200	42.9500	18.5700	0.0000	0.0500	22.4200	0.4700	8.2800
14	29.1400	0.1200	18.3700	46.1900	69.0800	1.7200	48.4100	0.0400	21.1800	1.1900	21.7400	1.9700	0.1000	15.7700	30.3900	1.0400	0.0200	2.0800	43.6000	47.7300	9.8000
15	22.2000	13.1300	22.6000	25.7500	4.9200	6.6800	9.1400	2.3900	17.9400	0.2100	20.5400	40.2500	26.5900	65.2000	47.9700	23.8800	0.0400	90.0700	35.6000	42.4700	54.4900
16	43.4000	0.0200	0.0000	0.0200	50.0100	0.0000	0.0200	0.8450	32.4100	0.0300	0.0000	21.6200	0.0000	0.2200	33.1100	6.8400	0.0000	23.3800	30.6200	12.0300	12.1100
17	0.0800	0.0000	0.0000	0.0000	0.0000	0.0100	0.0000	0.0100	0.0300	0.0000	0.0000	0.0000	0.0000	0.0000	0.0200	0.0000	0.0000	0.0000	0.0100	0.0000	16.4400
18	105.7200	6.5900	8.4700	19.9400	102.5000	2.6600	12.5500	17.9000	72.8900	0.0000	11.8100	7.7700	16.0400	13.1100	90.4600	2.4600	0.0100	20.8400	37.3500	31.9100	34.6200
19	30.1400	0.0400	21.7900	0.0500	87.9300	0.0000	0.0100	61.3600	31.1200	0.0000	15.6100	19.7400	11.6400	7.0600	20.2800	33.7000	1.8200	0.0000	24.5500	112.0600	12.0800
20	50.0300	0.0100	15.4700	0.0200	101.1800	0.0000	0.0100	59.4500	26.6200	0.0000	0.0400	2.6700	0.0500	0.0200	47.0700	0.0000	0.0000	43.0600	40.9800	13.0000	16.6800

21	6.7800	7.6300	13.1400	14.1000	21.5100	3.7000	13.5200	0.0900	18.4300	0.0900	0.7600	24.1600	23.4200	50.2200	2.1700	10.9300	0.0200	39.0800	43.7700	22.4700	0.0000
22	13.9400	0.0000	0.0000	0.0200	70.0700	0.0000	0.0100	0.0100	23.0500	0.0000	0.0000	0.0300	0.0000	0.0200	16.1400	0.0000	0.0000	0.1400	0.1300	0.0000	0.8000
23	30.6300	0.0200	0.0300	1.0000	33.0000	1.0500	0.0100	12.7000	26.9600	0.0000	0.4000	6.1300	0.0500	16.3000	20.1900	0.0800	0.0000	7.6900	12.4900	0.1900	1.0900
24	3.0900	0.0200	1.0000	1.2400	11.1700	0.0100	0.0000	0.0000	9.2300	0.0000	0.0000	0.0600	0.0300	0.0100	0.3800	0.2000	0.0000	0.0100	0.0000	4.1200	1.0000
25	5.6100	2.2200	1.2000	0.4300	7.3700	0.0500	0.1000	0.0000	4.2200	0.0000	0.3000	5.1900	3.6800	16.9000	4.8700	4.0400	0.0000	2.5000	5.2100	1.6000	0.3300
26	2.7300	0.0000	0.0000	0.0000	14.0400	0.0000	0.0000	0.0400	1.5600	0.0000	0.0000	0.0400	0.0200	0.0000	3.5400	0.0000	0.0000	0.0000	0.0000	0.0000	1.0500
27	115.0200	########	########	99.8300	52.4700	########	90.4900	75.0700	25.9600	24.2100	21.9300	########	########	44.2300	34.8400	########	13.3300	106.4900	297.8800	146.3000	23.6500