

中文赛题 C：预测Wordle结果

背景

Wordle是由《纽约时报》每天推出的一种受欢迎的益智游戏。玩家们需要在六次或更少的猜测中猜出一个由五个字母组成的单词，并在每次猜测后得到反馈。在这个版本中，每个猜测必须是英语中的一个实际单词。比赛中不被认可为单词的猜测是不允许的。Wordle在人们中不断增长的流行度中，现在有60多种语言的游戏版本可供选择。

《纽约时报》网站上关于Wordle的说明指出，在提交单词后，瓷砖的颜色会发生变化。黄色的瓷砖表示该瓷砖中的字母在单词中，但位置不正确。绿色的瓷砖表示该瓷砖中的字母在单词中，位置正确。灰色的瓷砖表示该瓷砖中的字母根本不包含在单词中（见附件2）。图1是一个示例解决方案，其中在三次尝试中找到了正确答案。

Wordle

A	D	I	E	U
A	V	O	I	D
A	P	H	I	D

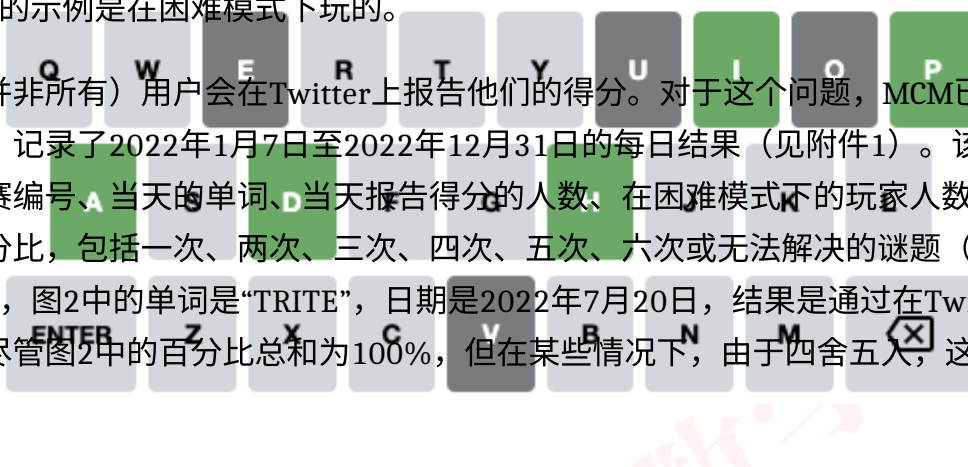


关注数学模型
获取更多资讯

Figure 1: Example Solution of Wordle Puzzle from July 21, 2022[3]

玩家可以在常规模式或“困难模式”下玩。Wordle的困难模式通过要求一旦玩家在单词中找到正确的字母（瓷砖为黄色或绿色），就必须在随后的猜测中使用这些字母来使游戏更加困难。图1中的示例是在困难模式下玩的。

许多（但并非所有）用户会在Twitter上报告他们的得分。对于这个问题，MCM已经生成了一个文件，记录了2022年1月7日至2022年12月31日的每日结果（见附件1）。该文件包括日期、比赛编号、当天的单词、当天报告得分的人数、在困难模式下的玩家人数，以及猜出单词的百分比，包括一次、两次、三次、四次、五次、六次或无法解决的谜题（表示为X）。例如，图2中的单词是“TRITE”，日期是2022年7月20日，结果是通过在Twitter上收集得到的。尽管图2中的百分比总和为100%，但在某些情况下，由于四舍五入，这可能不是真实的。



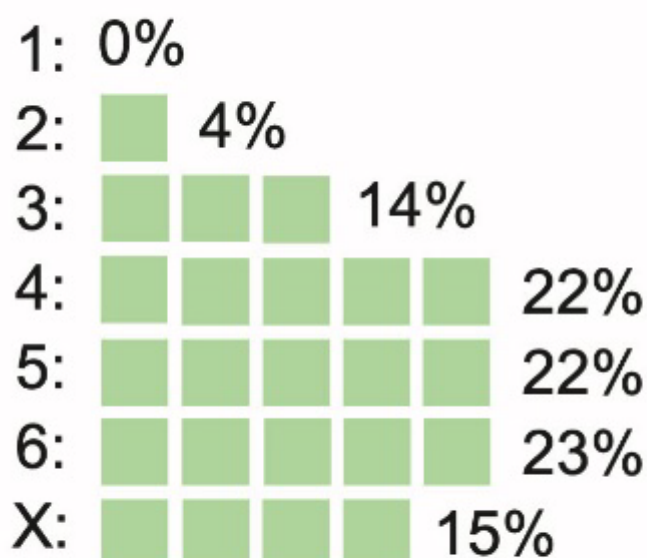
Wordle Stats @ WordleStats • 1h

🤖 Automated

#Wordle 396 2022-07-20

42,237 results found on Twitter.

3,685 hard mode players.



#Wordle396

Figure 2: Distribution of the Reported Results for July 20, 2022 to Twitter[4]



关注数学模型
获取更多资讯

要求

纽约时报要求您对该文件中的结果进行分析，以回答几个问题。

- 报告的结果数量每天都有所不同。开发一个模型来解释这种变化，并使用您的模型创建一个关于2023年3月1日报告结果数量的预测区间。是否有单词的属性会影响报告的得分中在困难模式下玩的比例？如果有，是怎样的？如果没有，为什么？
- 对于未来日期的给定解决方案单词，开发一个模型，使您可以预测报告结果的分布。换句话说，预测未来日期的相关百分比（1、2、3、4、5、6、X）的分布。您的模型和预测有哪些不确定性？请举一个关于2023年3月1日单词EERIE的预测的具体例子。您对您的模型的预测有多自信？
- 开发并总结一个模型，通过难度分类解决方案单词。确定与每个分类相关联的给定单词的属性。使用您的模型，单词EERIE有多难？讨论您的分类模型的准确性。
- 列出并描述该数据集的其他有趣特征。
- 最后，用一页至两页的信函，对纽约时报的谜题编辑总结您的结果。

您的PDF解决方案总页数不超过25页，其中包括：

- 一页摘要。
- 目录表。
- 您的完整解决方案。
- 一页至两页的信函。
- 参考文献列表。

注意：MCM竞赛有25页的限制。您的所有提交内容都计入25页限制（总结表、目录表、报告、参考文献列表以及任何附录）。您必须引用您报告中使用的想法、图片和其他材料的来源。



关注数学模型
获取更多资讯

How To Play

Guess the Wordle in 6 tries.

- Each guess must be a valid 5-letter word.
- The color of the tiles will change to show how close your guess was to the word.

Examples

W E A R Y

W is in the word and in the correct spot.

P **I** L L S

I is in the word but in the wrong spot.

V A G **U** E

U is not in the word in any spot.

术语表

纽约时报：一份总部位于美国纽约市的日报，以印刷和在线出版为主。 Twitter：一种社交网络网站，允许用户发布不超过 280 个字符的短消息（最初是 140 个字符）。 解决（Wordle 拼图）：按正确的顺序输入正确的字母以形成当天的 Wordle 单词。

参考资料

注：我们提供以下引文以支持问题陈述。我们从这些资源中提取了重要的想法。这些网站上没有解决MCM问题所需的其他信息。解决这个 MCM 问题不需要访问纽约时报或 Twitter 网站。

[1] Wordle logo from The New York Times website. Accessed on December 13, 2022 at <https://nytcassets.nytimes.com/2022/08/cropped-Screen-Shot-2022-08-24-at-8.49.39-AM.png>. [2] "Wordle-The New York Times." The New York Times, 2022. Accessed December 13, 2022 at <https://www.nytimes.com/games/wordle/index.html>. [3] "Wordle-The New York Times." The New York Times, July 21, 2022. [4] "Wordle Stats." Twitter, July 20, 2022.



关注数学模型
获取更多资讯

Problem C: Predicting Wordle Results

Background

Wordle is a popular puzzle currently offered daily by the New York Times. Players try to solve the puzzle by guessing a five-letter word in six tries or less, receiving feedback with every guess. For this version, each guess must be an actual word in English. Guesses that are not recognized as words by the contest are not allowed. Wordle continues to grow in popularity and versions of the game are now available in over 60 languages.

The New York Times website directions for Wordle state that the color of the tiles will change after you submit your word. A yellow tile indicates the letter in that tile is in the word, but it is in the wrong location. A green tile indicates that the letter in that tile is in the word and is in the correct location. A gray tile indicates that the letter in that tile is not included in the word at all (see Attachment 2)[2]. Figure 1 is an example solution where the correct result was found in three tries.



关注数学模型
获取更多资讯

Wordle

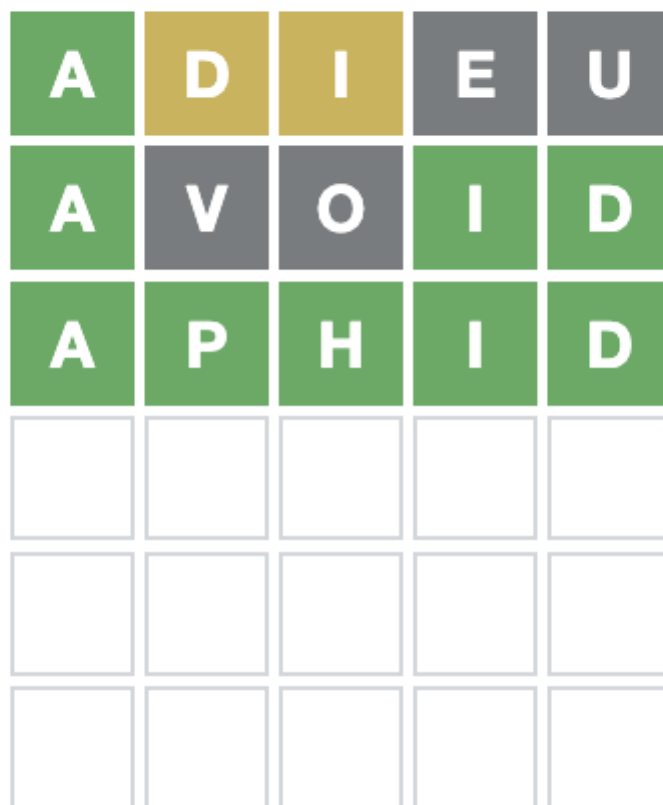


图 1: 2022年7月21日单词拼图的示例解决方案[3]

Players can play in regular mode or “Hard Mode.” Wordle’s Hard Mode makes the game more difficult by requiring that once a player has found a correct letter in a word (the tile turns yellow)



or green), those letters must be used in subsequent guesses. The example in Figure 1 was played in Hard Mode.

Many (but not all) users report their scores on Twitter. For this problem, MCM has generated a file of daily results for January 7, 2022 through December 31, 2022 (see Attachment 1). This file includes the date, contest number, word of the day, the number of people reporting scores that day, the number of players on hard mode, and the percentage that guessed the word in one try, two tries, three tries, four tries, five tries, six tries, or could not solve the puzzle (indicated by X). For example, in Figure 2 the word on July 20, 2022 was “TRITE” and the results were obtained by mining Twitter. Although the percentages in Figure 2 sum to 100%, in some cases this may not be true due to rounding.

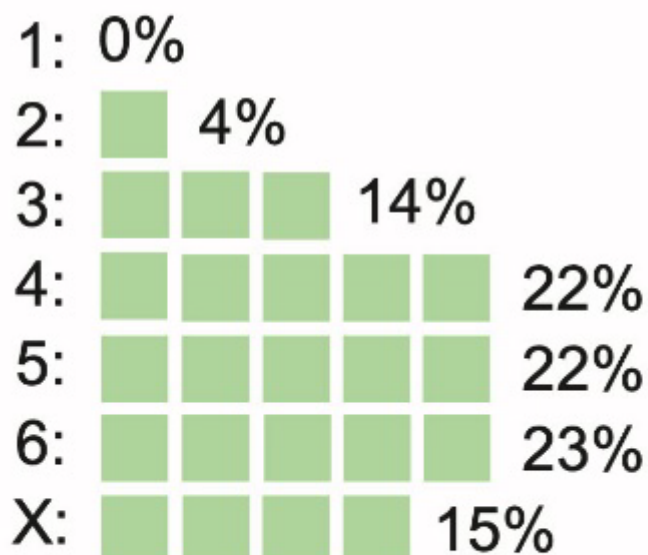
Wordle Stats @ WordleStats • 1h

🤖 Automated

#Wordle 396 2022-07-20

42,237 results found on Twitter.

3,685 hard mode players.



#Wordle396

图2:2022年7月20日报告结果在Twitter上的分布[4]

Requirement



关注数学模型
获取更多资讯

You have been asked by the New York Times to do an analysis of the results in this file to answer several questions.

- The number of reported results vary daily. Develop a model to explain this variation and use your model to create a prediction interval for the number of reported results on March 1, 2023. Do any attributes of the word affect the percentage of scores reported that were played in Hard Mode? If so, how? If not, why not?
- For a given future solution word on a future date, develop a model that allows you to predict the distribution of the reported results. In other words, to predict the associated percentages of (1, 2, 3, 4, 5, 6, X) for a future date. What uncertainties are associated with your model and predictions? Give a specific example of your prediction for the word EERIE on March 1, 2023. How confident are you in your model's prediction?
- Develop and summarize a model to classify solution words by difficulty. Identify the attributes of a given word that are associated with each classification. Using your model, how difficult is the word EERIE? Discuss the accuracy of your classification model.
- List and describe some other interesting features of this data set.

Finally, summarize your results in a one- to two-page letter to the Puzzle Editor of the New York Times.

Your PDF solution of no more than 25 total pages should include:

- One-page Summary Sheet.
- Table of Contents.
- Your complete solution.
- One- to two-page letter.
- Reference List.

Note: The MCM Contest has a 25-page limit. All aspects of your submission count toward the 25-page limit (Summary Sheet, Table of Contents, Report, Reference List, and any Appendices). You must cite the sources for your ideas, images, and any other materials used in your report.

Attachments

1.Data File. Problem C Data Wordle.xlsx

THE ATTACHED DATA FILE CONTAINS THE ONLY DATA YOU SHOULD USE FOR THIS PROBLEM. All information needed for this problem is given in the problem statement and the



关注数学模型
获取更多资讯

data file. You do not need to visit the New York Times website nor Twitter website. There is no additional information to be found on these sites.

Data File Entry Descriptions

- Date: The date in mm-dd-yyyy (month-day-year) format of a given Wordle puzzle.
- Contest number: An index of the Wordle puzzles, beginning with 202 on January 7, 2022.
- Word: The solution word players are trying to guess on the associated date and contest number.
- Number of reported results: The total number scores that were recorded on Twitter that day.
- Number in hard mode: The number of scores on Hard mode recorded on Twitter that day.
- 1 try: The percentage of players solving the puzzle in one guess.
- 2 tries: The percentage of players solving the puzzle in two guesses.
- 3 tries: The percentage of players solving the puzzle in three guesses.
- 4 tries: The percentage of players solving the puzzle in four guesses.
- 5 tries: The percentage of players solving the puzzle in five guesses.
- 6 tries: The percentage of players solving the puzzle in six guesses.
- 7 or more tries (X): The percentage of players that could not solve the puzzle in six or fewer tries. Note: the percentages may not always sum to 100% due to rounding.

2.Directions of Wordle posted to the New York Times website.[2]



关注数学模型
获取更多资讯

How To Play

Guess the Wordle in 6 tries.

- Each guess must be a valid 5-letter word.
- The color of the tiles will change to show how close your guess was to the word.

Examples

W **E** **A** **R** **Y**

W is in the word and in the correct spot.

P **I** **L** **L** **S**

I is in the word but in the wrong spot.

V **A** **G** **U** **E**

U is not in the word in any spot.

Glossary

New York Times: A daily newspaper based in New York City, New York, USA published in print and online.

Twitter: A social networking site that allows users to broadcast short posts of no more than 280 characters (increased from initial 140 characters).

Solve (the Wordle puzzle): Enter the correct letters in the correct order to form the Wordle word of the day.

References

Note: We provide the following citations to support the Problem Statement. We have pulled the important ideas from these resources. There is no additional information on these websites needed to solve this MCM problem. Access to the New York Times or Twitter website is not required to solve this problem.



关注数学模型
获取更多资讯

[1] Wordle logo from The New York Times website. Accessed on December 13, 2022 at <https://nytco-assets.nytimes.com/2022/08/cropped-Screen-Shot-2022-08-24-at-8.49.39-AM.png>.

[2] “Wordle-The New York Times.” The New York Times, 2022. Accessed December 13, 2022 at <https://www.nytimes.com/games/wordle/index.html>.

[3] “Wordle-The New York Times.” The New York Times, July 21, 2022.

[4] “Wordle Stats.” Twitter, July 20, 2022.

数学模型
微信公众号: MATHmodels



关注数学模型
获取更多资讯