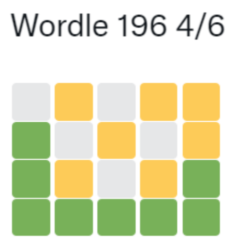
4.1: Discrete Random Variables

Wordle is a popular puzzle game that was created by the New York Times in October of 2021. To play the game, a player tries to guess the five letter word of the day. Each player has a total of six tries to guess the word. After each guess, the colors of tiles will change to indicate how close the guess was to the word. If a tile turns green, the letter is in the correct spot. If a tile turns yellow, the letter is in the word but not in the correct spot. If a letter turns gray, it is not a letter in the word. According to twitter, on April 28, 2022, 98,967 individuals played Wordle.



Adapted from "Wordle Emoji Screenshot" by Levi OP is licensed under CC BY-SA 4.0.

The following table sorts the number of people who played by the number of guesses taken to solve the puzzle.

1. Complete the table by computing the relative frequencies (rounded to three decimal places) for each value of x. Recall that n is the total number of people in the table.

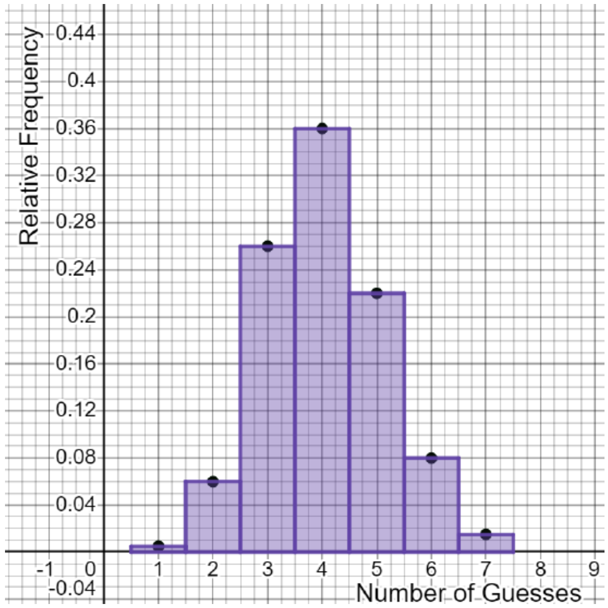
Number of Guesses, x_i	**Number of people, f_i**	**Relative Frequency, f_i/n**
1	494	
2	5938	
3	25731	
4	35628	
5	21772	
6	7917	
More than 6	1487	

The number of guesses is an example of a **discrete random variable**. We can list the values of a discrete random variable in order. In other words, we can *count* the values of the variable. In this example, the variable x can take on the values {1, 2, 3, 4, 5, 6, 7} (if an individual takes 7 guesses, they have lost the game because they have run out of tries). If we randomly select a game from this population, the probability that it was won in a certain number of tries is equal to the relative frequency of the chosen number of guesses.

A **discrete probability distribution** assigns a probability to all possible values of a discrete random variable. Each probability is between 0 and 1, and the sum of all probabilities is 1.

Finding Probability from a Discrete Probability Distribution

The discrete probability distribution in the Wordle example above can be represented with a histogram.



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

2. What is the shape of the histogram? What does the shape suggest to you about the number of guesses taken to win Wordle?

3. Imagine that we randomly select a Wordle game. Use the probability distribution above to answer the following questions.

a. Find the probability that the game is won in more than 5 guesses. Using probability notation, we would denote this probability as $P(x > 5)$. (Hint: we are asking what the probability is that the word was guessed on the sixth try OR the word was not guessed).

b. Find the probability that the game is won in at most 4 guesses. Use probability notation in your answer.

c. Find the probability that the game is won between 3 and 6 guesses. Use probability notation in your answer.

d. Find the probability that the game is won (in other words, the number of guesses is *not* 7). Use probability notation in your answer.

e. Three friends are competing with each other through Wordle. What is the probability that all three individuals guess in less than 3 tries? Assume guesses are independent. Use probability notation in your answer. Round your answer to five decimal places.

f. You want to know how you compare to other Wordle players. You want to know the average number of guesses it takes to win Wordle. Use the histogram to estimate the mean number of guesses.

Finding the Mean of a Discrete Probability Distribution

The mean of a probability distribution is the mean of a distribution of relative frequencies. If the probabilities are exact, the mean is a **population mean**. We have seen that the symbol used to denote a population mean is μ (read as "mew") and the symbol used to denote a sample mean is $\bar{x}$ (read as "x bar"). In a previous exercise, you were tasked with computing a mean from a frequency distribution table. The formula used was

$$\text{mean } = \frac{\sum x_i \cdot f_i}{n}$$

where $n = \sum f_i$. We can use algebra to write an equivalent equation,

$$\text{mean } = \sum \frac{f_i}{n} \cdot x_i$$

Noting that f_i / n is a relative frequency, and knowing that relative frequencies are regarded as probabilities in a discrete probability distribution, we can say,

$$\text{population mean } = \mu = \sum x_i \cdot P\left(x_i\right)$$

In summary, the **mean** or **expected value** (denoted $E(x)$) **of a discrete probability distribution** is found using

$$\mu = E(x) = \sum x_i \cdot P\left(x_i\right).$$

The mean of a probability distribution represents both the average value of the random variable and the center of the distribution.

4. Compute the expected number of guesses it takes to win Wordle. Compare this to the estimation you made in 3f.

Finding the Standard Deviation of a Discrete Probability Distribution

The population standard deviation of a probability distribution is denoted by the Greek letter σ (read "sigma"). It is computed by adding all square deviations, dividing by the total frequencies, and square rooting. Using algebra, we can manipulate the inside of the square root:

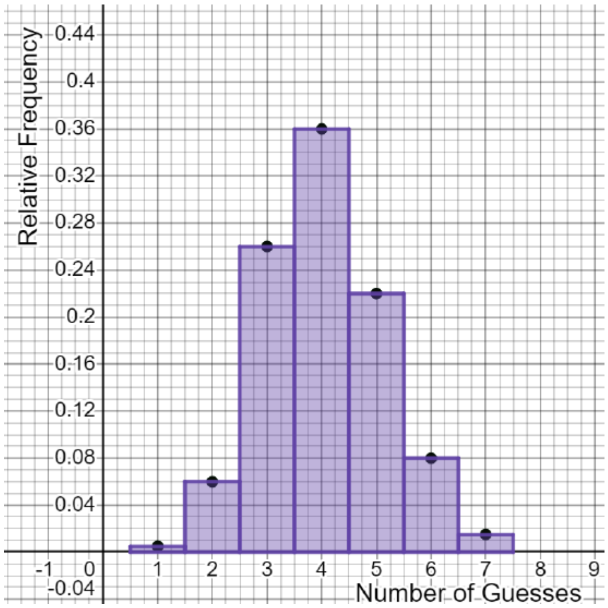
$$\begin{aligned}\text{population standard deviation } &= \sqrt{\frac{\Sigma\left(x_i - \mu\right)^2}{n}} = \sqrt{\frac{f_1\left(x_1 - \mu\right)^2 + f_2\left(x_2 - \mu\right)^2 + \ldots + f_n\left(x_n - \mu\right)^2}{n}} \\ &= \sqrt{\sum \frac{f_i}{n} \cdot \left(x_i - \mu\right)^2}\end{aligned}$$

where $n = \sum f_i$. Notice that f_i / n is the same as the probability. Therefore, we can find the standard deviation of a probability distribution using the following formula:

$$\text{population standard deviation } = \sigma = \sqrt{\sum P\left(x_i\right) \cdot \left(x_i - \mu\right)^2}$$

5. For the number of guesses in the game of Wordle, the population standard deviation is $\sigma = 1.1$ guesses. Interpret the standard deviation in context.

Area and Probability



Images are created with the graphing calculator, used with permission from Desmos Studio PBC.

6. In the Wordle histogram, the width of each bar is 1. Recall, the area of a rectangle is found using the width times the length.
   a. What is the area of the bar centered at 1?

   b. What is the area of the bar centered at 2?

   c. What is the relationship between a value's area and its probability?

   d. What is the total area of all the bars?

Apply it!

7. Finn is purchasing a new car from Ford. Ford offers buyers a maintenance plan which will cost an additional $35 per month. It will cover oil changes, routine maintenance, and any major repairs. The maintenance plan provides coverage for the first 6 years of ownership. On average, Finn drives 10,000 miles per year. He will get his car serviced every 5,000 miles. The average cost of service is $150.

   a. What is the cost of services for the first six years if Finn does not purchase the maintenance plan?

   b. Below is a probability distribution where you can see the potential cost of routine maintenance and possible major repairs in the first six years of car ownership. Standard services are not included in the costs. What is the expected cost of routine maintenance and possible major repairs for the first 6 years?

Annual Maintenance and Repair Costs for the First Six Years, \(x\)	0	$500	$1,000	$1,500	$2,000	$2,500
Probability, \(P(x)\)	0.02	0.52	0.21	0.14	0.08	0.03

   c. Compute the total expected cost for services, routine maintenance, and repairs for the first six years.

   d. What would Finn pay if he purchases the maintenance plan?

   e. Should Finn purchase the maintenance plan? Support your answer.