
IS IT NORMAL: A NOVEL APPROACH OF UTILIZING PYTHON 3.6 TO GENERATE A NORMAL GAUSSIAN DISTRIBUTION

A PREPRINT

Deepak Ramalingam
High School Senior
Monta Vista High School
Cupertino, California 95014
rdeepak2002@gmail.com

Kirtan U. Shah
High School Senior
Monta Vista High School
Cupertino, California 95014
kirtan.u.shah@gmail.com

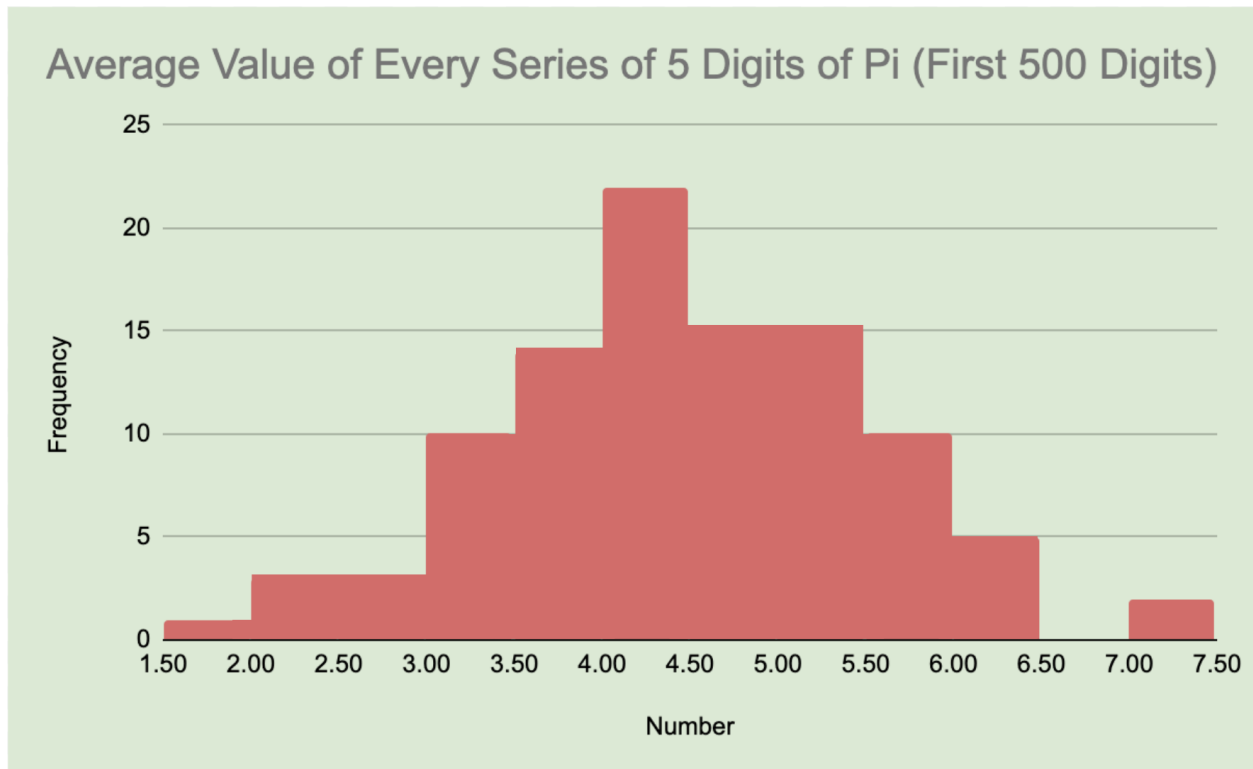
September 9, 2019

ABSTRACT

The following paper describes the manner in which we generated a normal gaussian distribution by utilizing the Python 3.6 programming language in order to create a program, program.py, to read in a text file containing the first one million digits of pi and output the average of every 5 digits to a new text file, then using the Google Sheets web application in order to create a visual histogram using the outputted data.

1 Histogram

Figure 1



2 Introduction

We received the first one million digits of pi as a text file from <https://www.angio.net/pi/digits.html>, then we used the Python 3.6 programming language to create the code seen in figure 2. This code reads the first million digits of pi as from the text file, then uses a for loop in order to average every 5 digits of pi. We chose this as our data because we are interested in relating statistics to the irrational number pi, which is used in wave functions and calculating circular properties.

Figure 2

```
1 limit = 60000
2 avgNum = 5
3
4 f = open("pimil.txt")
5
6 pi = f.read()
7 pi = pi.replace("\n", "")
8 pi = pi.replace(" ", "")
9
10 f2 = open("output.txt", "r+")
11
12 avg = 0.0
13
14 for i in range(0, limit, avgNum):
15     substr = pi[i:i+avgNum]
16     for j in range(avgNum):
17         avg += int(substr[j])
18     avg = avg / avgNum
19     f2.write(str(avg) + "\n")
20     avg = 0.0
```

3 Description

Minimum	First Quartile	Median	Third Quartile	Maximum	Mean	Standard Deviation
1.60	3.60	4.40	5.20	7.20	4.42	1.09

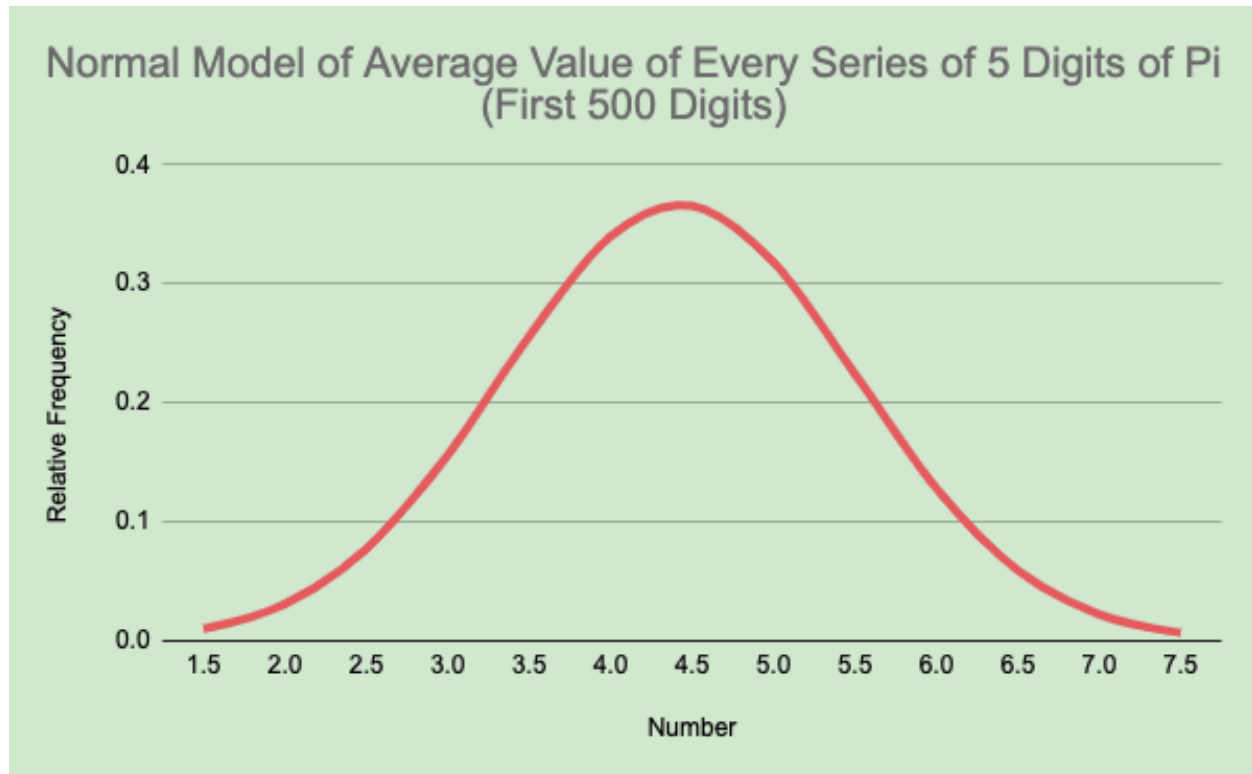
The shape of the data is unimodal and approximately symmetric with no skew. The mean is 4.42, the median is 4.40, and the mode is 4.00. The standard deviation is 1.09, the interquartile range is 1.60, and the range is 5.60. There are no outliers, but there is a gap between 6.5 and 7.

In my distribution it is better to use mean and standard deviation because the mean reveals that 4.42 is the average digit since 4.5 it is the average value of the digits from 0 to 9. The standard deviation of 1.09 also reveals that the data is approximately a normal distribution since a normal distribution has a standard deviation of 1.

4 Normal Model

The Normal model, $N(4.42, 1.09)$ is shown in figure 3. The graph is centered at the mean of our data and has the same scale as the histogram from figure 1. The 68-95-99.7% rule can clearly be shown in the graph.

Figure 3



5 Comparison of Distribution to Normal Model

68.28% (1707/2500) of our data is within 1 standard deviation of the mean. 95.45% (1909/2000) of our data is within 2 standard deviations of the mean. 99.73% (374/375) of our data is within 3 standard deviations of the mean. These percentages are approximately equal to 68%, 95%, and 99.7%. The original distribution has an approximately symmetrical distribution just like the Normal Model. Thus, the Normal Model is useful because the original distribution is similar to the Normal Model which allows us to make conclusions from the original data based off the Normal Model.

References

“Digits of Pi - Up to 1 Million Digits.” Digits of Pi - Up to 1 Million Digits, Angio Dot Net, 3 Jan. 2016, 5:59:22, www.angio.net/pi/digits.html.