cheg304 hw5 question 2

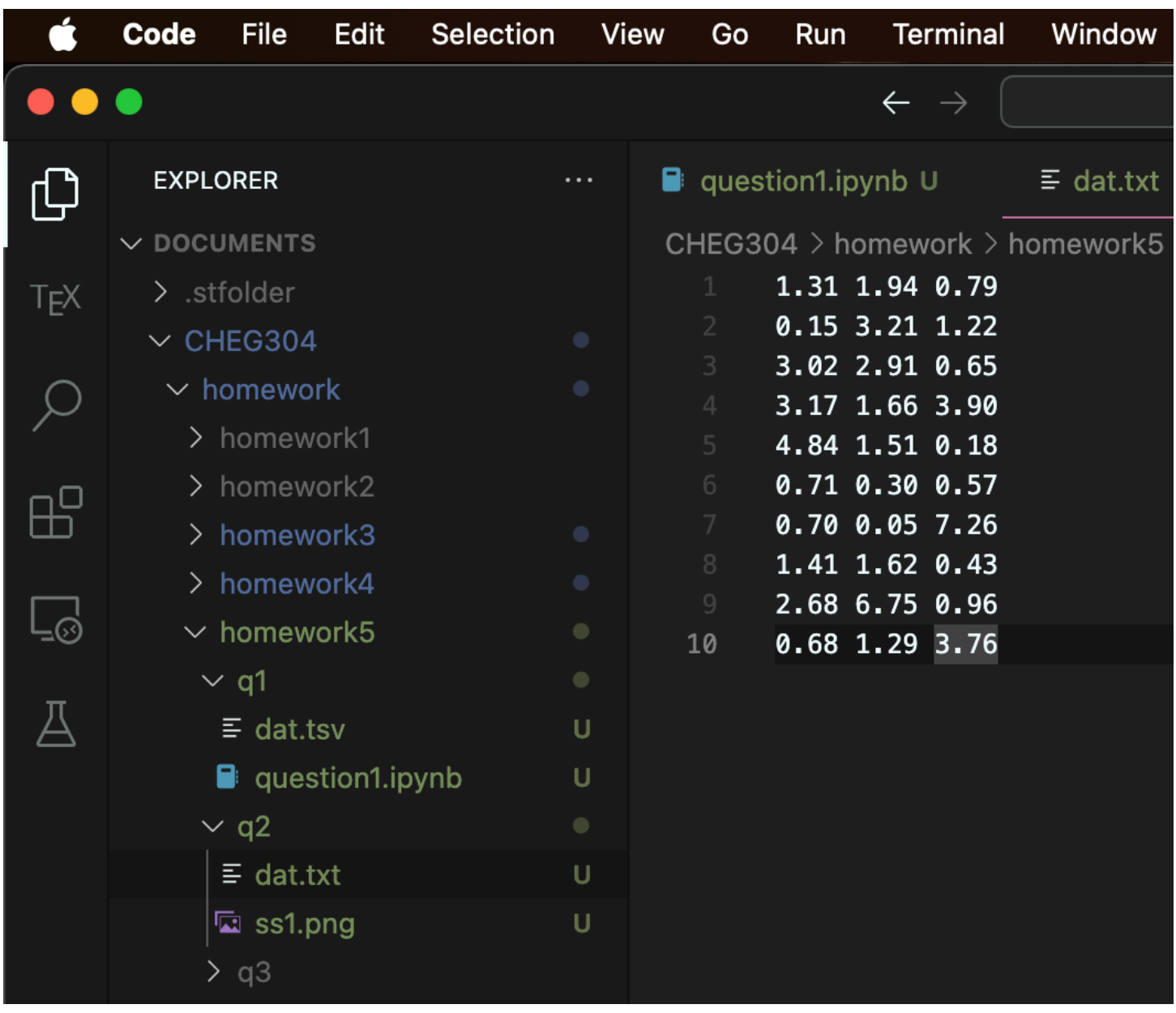
AUTHOR
k.wodehouse

PUBLISHED
March 12, 2025

the first step of finding the mean and such for this data was copying it into a .txt file and bringing that into my IDE.

1.31	1.94	0.79
0.15	3.21	1.22
3.02	2.91	0.65
3.17	1.66	3.90
4.84	1.51	0.18
0.71	0.30	0.57
0.70	0.05	7.26
1.41	1.62	0.43
2.68	6.75	0.96
0.68	1.29	3.76

- Show your work (it's fine to use a computer program's built-in functions) as to how you found the mean, median, and variance for this sample data.



it would be TERRIBLE to manually move around this data and is really easy to just copy paste and use an extra line of code. thank you for coming to my ted talk.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

df = pd.read_csv('dat.txt', sep=' ', header=None)
data = np.array(  pd.concat([df[0], df[1], df[2]])  )
data
```

```
array([1.31, 0.15, 3.02, 3.17, 4.84, 0.71, 0.7 , 1.41, 2.68, 0.68, 1.94,
       3.21, 2.91, 1.66, 1.51, 0.3 , 0.05, 1.62, 6.75, 1.29, 0.79, 1.22,
       0.65, 3.9 , 0.18, 0.57, 7.26, 0.43, 0.96, 3.76])
```

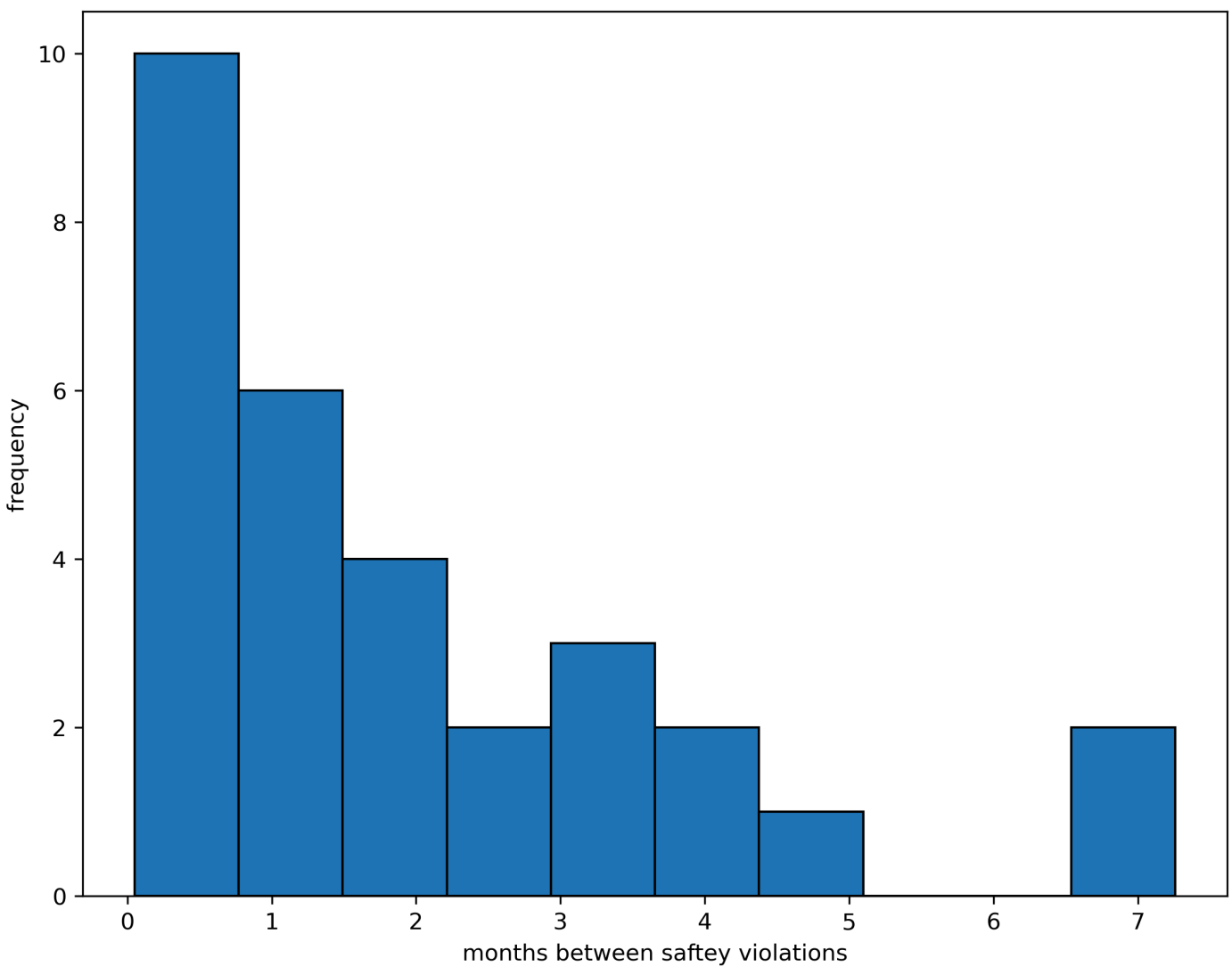
now for mean median and variance (using some delicious numpy functions #numpy)

```
print(f'mean: {data.mean():.2f}')
print(f'median: {np.median(data):.2f}')
print(f'variance: {data.var():.2f}')
```

```
mean: 1.99
median: 1.36
variance: 3.33
```

i like to imagine that saftey incidents are rare events, and our distribution for time between rare events (time between poisson events) is the exponential distribution.

```
fig,ax = plt.subplots(figsize=(9,7), dpi=300)
ax.hist(data, edgecolor='black')
ax.set(xlabel='months between saftey violations', ylabel='frequency');
```



and the plot supports the idea that this would be our exponential distribution!

now calculating the probability is as follows using the pdf of our exponential rv

$$
\begin{aligned}
P(X > 2) &= \int_2^\infty \frac{1}{2}\exp\frac{-x}{2} \\
&= \left[-\exp\frac{-x}{2}\right]_2^\infty \\
&= \exp(-1) \\
&\approx 0.368
\end{aligned}
$$

```
emperical = len(data[data>2]) / len(data)
print(f'proportion of data > 2 months: {emperical:.3f}')
```

```
proportion of data > 2 months: 0.333
```

the theoretical 0.368 seems pretty compatible to the 0.333 to me. i would say compatible since the graph looks pretttty exponential decay to me and the theoretical proportion is just over 1/3 and the empirical propertion is 1/3.

```
# fillllller
```