

ENG 346

Data Structures and

Algorithms for Artificial

Intelligence

Statistical Operations and Numpy

Dr. Mehmet PEKMEZCİ

mpekmezci@gtu.edu.tr

<https://github.com/mehmetpekmezci/GTU-ENG-346>

Statistics

- Statistics is the discipline that concerns the collection, organization, analysis, interpretation, and presentation of data.
- **Population:** The entire group one desires information about. **Universal Set.**
- **Sample:** A subset of population taken because the entire population is usually too large to analyze. Try to represent all the characteristics of the population (As much as you can).
- **Mean:** The average
- **Median:** The mid point.
- **Mode:** The most common.

- **Variance:** Variance is a measure of how far the observed values in a dataset fall from the arithmetic mean, and is therefore a measure of spread - more specifically, it is a measure of variability.
- **Standard Deviation: (Square Root of Variance)** .Standard deviation is a measure of spread - more specifically, it is a measure of dispersion (or, the measure of variability!).
- **Standard Error : (you can see how well your sample data represents your population.)** Standard error is another measure of spread. The most common standard error is the standard error of the mean, and used to measure sampling error as it measures how accurately the mean of a sample distribution represents the mean of the population.

Mean-Median-Mode

Mean: The average

Median: The mid point.

Mode: The most common.

Example :

Speed = [99,86,87,88,111,86,103,87,94,78,77,85,86]

Mean = $(99+\dots+86)/13=89.77$

Median = 87 (When **sorted** , 87 is in the middle)

Mode = 86 (There are three 86 in the list)

```
import numpy as np
from scipy import stats as st
```

```
speed= [99,86,87,88,111,86,103,87,94,78,77,85,86]
```

```
mean=np.mean(speed)
median=np.median(speed)
mode=st.mode(speed)
```

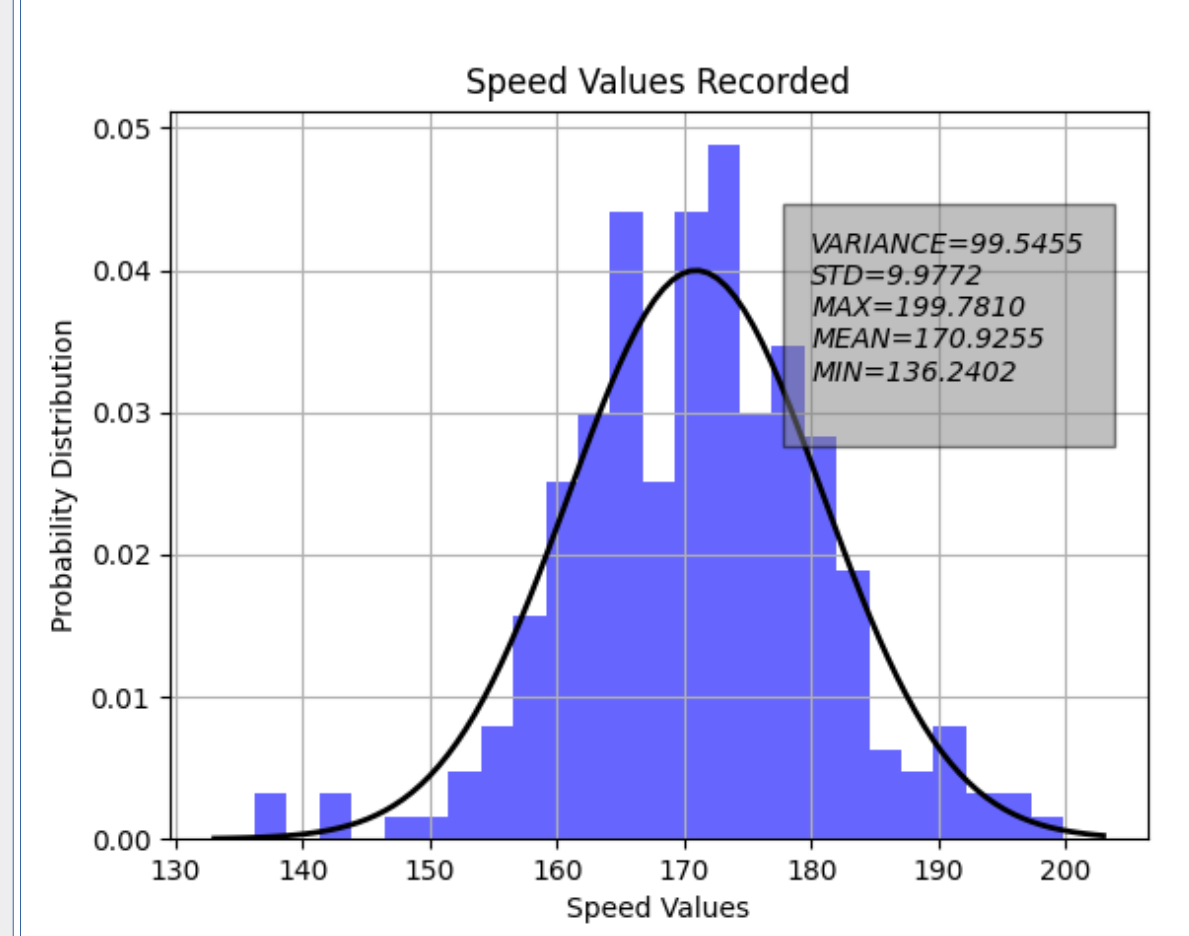
```
print(f"Mean={mean}")
print(f"Median={median}")
print(f"Mode={mode[0]}, appears {mode[1]} times in the list.")
```

```
mpekmezci@cobalt:~$ python3 01.mean-median-mode.py
Mean=89.76923076923077
Median=87.0
Mode=86, appears 3 times in the list.
mpekmezci@cobalt:~$
```

Variance / Standard Deviation

Variance: Variance is a measure of how far the observed values in a dataset fall from the arithmetic mean, and is therefore a measure of spread - more specifically, it is a measure of variability.

```
import numpy as np
from scipy.stats import norm
import matplotlib.pyplot as plt
data = np.random.normal(170, 10, 250) # 170 mean, 10 std, 250 number of samples
variance=np.var(data)
max_val=np.max(data)
min_val=np.min(data)
mu, std = norm.fit(data) # Fit a normal distribution to the data
plt.hist(data, bins=25, density=True, alpha=0.6, color='b')
# Plot the PDF.
xmin, xmax = plt.xlim()
x = np.linspace(xmin, xmax, 100)
p = norm.pdf(x, mu, std)
plt.plot(x, p, 'k', linewidth=2)
plt.title("Speed Values Recorded")
plt.grid(True)
plt.text( 180,0.03, f"VARIANCE={float(variance):.4f}\n STD={float(std):.4f}\n"+
          f"MAX={float(max_val):.4f}\n MEAN={float(mu):.4f}\n"
          f"MIN={float(min_val):.4f}\n",
          style='italic',  bbox={ 'facecolor': 'gray','alpha': 0.5, 'pad': 10 })
plt.xlabel("Speed Values")
plt.ylabel("Probability Distribution")
plt.show()
```



```
$ vi 02.variance.py
$ python3 02.variance.py
```