Mathematics_HW01_Assignment

October 16, 2025

FA25: ISE-201 Sec 33 - Math Dec. and Data Science

HW01 Assigment

Name: Bishakha

SJSU ID: 019178354

Date: Oct 16 2025

Question 6.1.4: For any set of data values, is it possible for the sample standard deviation to be larger than the sample mean? If so, give an example.

Answer: yes, it is possible , when values in dataset are at extreme.

Fo eg: [1,2,30) Mean = 33/3=11

st variance = (100+81+361)/2=542/2=271

standard deviation = sqrt(271)=16.46

std 16.46 is greater than mean 11.

Question 6.1.5: Can the sample standard deviation be equal to zero? If so, give an example.

Answer:

It is only possible when all the datapoints are same, in which case the mean will the same number/datapoint. so then the distance between the mean and data points will be 0 and so the variance will be zero and hence the standard deviation will be 0 too.

But it is very unlikely and un-natural to have a dataset like that.

so , ideally the standard deviation should not be equal to 0.

Question 6.1.14: In the 2000 Sydney Olympics, a special program initiated by the IOC president allowed developing countries to send athletes to the Olympics without the usual qualifying procedure. Here are the 71 times for the first round of the 100-meter men's swim (in seconds).

a. Find the sample mean and sample standard deviation of these 100-meter swim times.

b. Construct a dot diagram of the data.

c. Comment on anything unusual that you see.

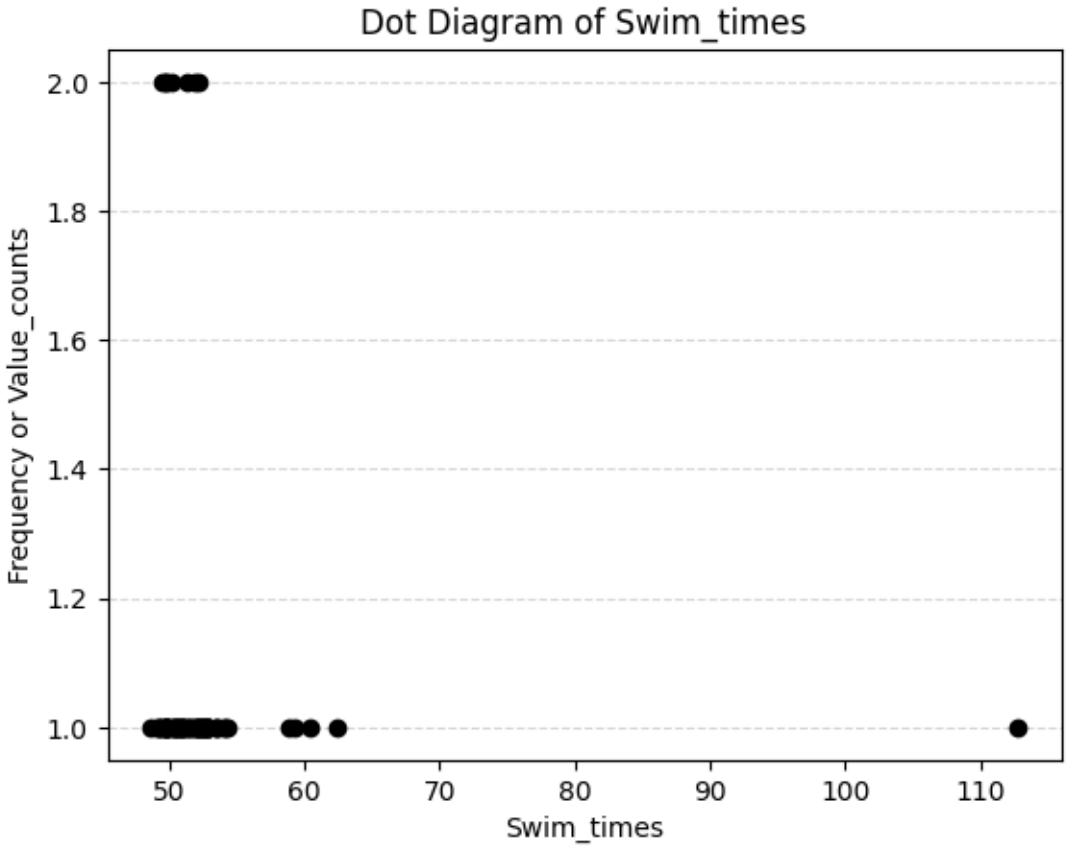
[ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
```

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

data = pd.read_csv('/content/drive/MyDrive/HW01Data.csv')
dataset = data['EX 6.1.14'].dropna()
dataset_arr = dataset.to_numpy()
print(dataset_arr)

### a. Find the sample mean and sample standard deviation of these 100-meter␣
 ↪swim times.

print(f"\nMean for 100 mt swim times is {np.mean(dataset_arr):.2f}")
print(f"\nstandard deviation for 100-mt swim times is {np.std(dataset_arr):.2f}␣
 ↪")

### b. Construct a dot diagram of the data.

swim_times = data['EX 6.1.14'].dropna().astype(float)
value_counts = swim_times.value_counts().sort_index()

# Plot the dot diagram
for value, count in value_counts.items():
    plt.plot([value] * count, range(1, count + 1), 'ko')

plt.xlabel('Swim_times')
plt.ylabel('Frequency or Value_counts')
plt.title('Dot Diagram of Swim_times')
plt.grid(True, axis='y', linestyle='--', alpha=0.5)
plt.show()
```

```
[ 60.39  49.93  53.4   51.82  50.46  51.34  50.28  50.19  52.14  50.56
  52.72  50.95  49.74  49.16  52.57  52.53  52.09  52.4   49.75  54.06
  53.5   50.63  51.93  51.62  52.58  53.55  51.07  49.76  49.73  50.9
  59.26  49.29  52.78 112.72  49.79  49.83  52.43  51.28  52.22  49.76
  49.7   52.9   50.19  54.33  62.45  51.93  52.24  52.82  50.96  48.64
  51.11  50.87  52.18  54.12  50.49  49.84  52.91  52.52  50.32  51.52
  52.    52.85  52.24  49.45  51.28  49.09  58.79  49.74  49.32  50.62
  49.45]
```

```
Mean for 100 mt swim times is 52.65

standard deviation for 100-mt swim times is 7.60
```



Question c. Comment on anything unusual that you see.

Answer: While most people took 49-65 seconds swim time for 100-mts , 1 person took around 112 seconds.

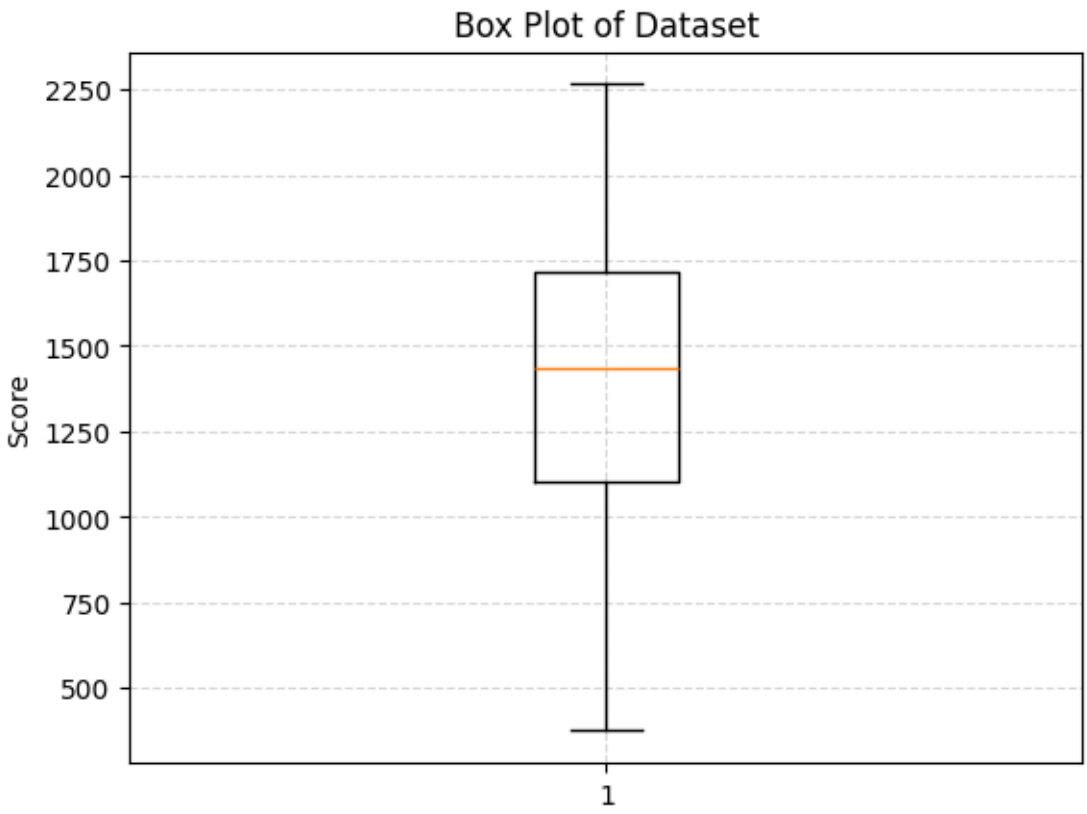
Question 6.2.5: The following data are the numbers of cycles to failure of aluminum test coupons subjected to repeated alternating stress at 21,000 psi, 18 cycles per second.

Construct a stem-and-leaf display for these data. Calculate the median and quartiles of these data. Does it appear likely that a coupon will "survive" beyond 2000 cycles? Justify your answer.

[ ]:
```
import pandas as pd
import numpy as np


data = pd.read_csv('/content/drive/MyDrive/HW01Data.csv')
```

```
dataset = data['EX 6.2.5'].dropna()
dataset_arr = dataset.to_numpy()
print(dataset_arr)
print(f"\nMedian for given data points is {np.median(dataset_arr):.2f}")
print(f"\nMode is {dataset.mode().to_list()}")
print(f"\nQuartile 1 is {np.quantile(dataset_arr, 0.25):.2f}")
print(f"\nQuartile 2 is {np.quantile(dataset_arr, 0.50):.2f}")
print(f"\nQuartile 3 is {np.quantile(dataset_arr, 0.75):.2f}")
print(f"\nQuartile 4 is {np.quantile(dataset_arr, 1):.2f}")
print("\n")
plt.boxplot(dataset)
plt.title('Box Plot of Dataset')
plt.ylabel('Score')
plt.grid(True, linestyle='--', alpha=0.5)
plt.show()
```

```
[1115. 1310. 1540. 1502. 1258. 1315. 1085.  798. 1020.  865. 2130. 1421.
 1109. 1481. 1567. 1883. 1203. 1270. 1015.  845. 1674. 1016. 1102. 1605.
  706. 2215.  785.  885. 1223.  375. 2265. 1910. 1018. 1452. 1890. 2100.
 1594. 2023. 1315. 1269. 1260. 1888. 1782. 1522. 1792. 1000. 1820. 1940.
 1120.  910. 1730. 1102. 1578.  758. 1416. 1560. 1055. 1764. 1330. 1608.
 1535. 1781. 1750. 1501. 1238.  990. 1468. 1512. 1750. 1642.]

Median for given data points is 1436.50

Mode is [1102.0, 1315.0, 1750.0]

Quartile 1 is 1102.00

Quartile 2 is 1436.50

Quartile 3 is 1716.00

Quartile 4 is 2265.00
```



.

Q. Does it appear likely that a coupon will "survive" beyond 2000 cycles? Justify your answer.

Ans: No, it is less likely that a coupon will survive beyond 2000 cycles, because most of data points lies between quartile 1,2 and 3. very few data points for more than 2000 cycles , which suggests that it is very likely to survive beyond 2000 cycles.

[ ]:

Question 6.2.7: Calculate the sample median, mode, and mean of the data in Exercise 6.2.5. Explain how these three measures of location describe different features in the data.

[ ]:
```
import pandas as pd
import numpy as np


data = pd.read_csv('/content/drive/MyDrive/HW01Data.csv')
dataset = data['EX 6.2.5'].dropna()
dataset_arr = dataset.to_numpy()
print(dataset_arr)
```

```
print(f"\nMedian is {np.median(dataset_arr):.2f}")
print(f"\nMode is {dataset.mode().to_list()}")
print(f"\nMean is {np.mean(dataset_arr):.2f}")
```

```
[1115. 1310. 1540. 1502. 1258. 1315. 1085.  798. 1020.  865. 2130. 1421.
 1109. 1481. 1567. 1883. 1203. 1270. 1015.  845. 1674. 1016. 1102. 1605.
  706. 2215.  785.  885. 1223.  375. 2265. 1910. 1018. 1452. 1890. 2100.
 1594. 2023. 1315. 1269. 1260. 1888. 1782. 1522. 1792. 1000. 1820. 1940.
 1120.  910. 1730. 1102. 1578.  758. 1416. 1560. 1055. 1764. 1330. 1608.
 1535. 1781. 1750. 1501. 1238.  990. 1468. 1512. 1750. 1642.]

Median is 1436.50

Mode is [1102.0, 1315.0, 1750.0]

Mean is 1403.66
```

Question 6.2.8: The female students in an undergraduate engineering core course at ASU self-reported their heights to the nearest inch. Construct a stem-and-leaf diagram for the height data and comment on any important features that you notice. Calculate the sample mean, the sample standard deviation, and the sample median of height.

[ ]:
```
import pandas as pd
import numpy as np


data = pd.read_csv('/content/drive/MyDrive/HW01Data.csv')
dataset = data['EX 6.2.8'].dropna()
dataset_arr = dataset.to_numpy()
print(dataset_arr)
print(f"\nMedian is {np.median(dataset_arr):.2f}")
print(f"\nstandard deviation is {np.std(dataset_arr):.2f}")
print(f"\nMean is {np.mean(dataset_arr):.2f}")
```

```
[62. 64. 66. 67. 65. 68. 61. 65. 67. 65. 64. 63. 67. 68. 64. 66. 68. 69.
 65. 67. 62. 66. 68. 67. 66. 65. 69. 65. 70. 65. 67. 68. 65. 63. 64. 67.
 67.]

Median is 66.00

standard deviation is 2.08

Mean is 65.81
```