



San Francisco Bay University

CS483L - Artificial Intelligence & Machine Learning Lab

2022 Summer Quiz#2

Instruction:

A. Put your answer right after each question in the answer sheet

1. Write a program to implement Bootstrapping algorithm to find the median number in the given dataset shown in the file “Bootstrapping example”

**Notice that it is suggested to create 10 new sets randomly*

Source code:

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

my_samples = []
sample = [12, 34, 45, 78, 99]
for i in range(10000):
    x = np.random.choice(sample, size=5, replace=True)
    my_samples.append(statistics.median(x))

freq_median = {12: 0, 34: 0, 45: 0, 78: 0, 99: 0}
for i in range(10000):
    if my_samples[i] == 12: freq_median[12] += 1
    elif my_samples[i] == 34: freq_median[34] += 1
    elif my_samples[i] == 45: freq_median[45] += 1
    elif my_samples[i] == 78: freq_median[78] += 1
    else: freq_median[99] += 1

key = freq_median.keys()
value = freq_median.values()
# create a table with key and value
table = pd.DataFrame({'key': key, 'Freq. of Median': value})
```

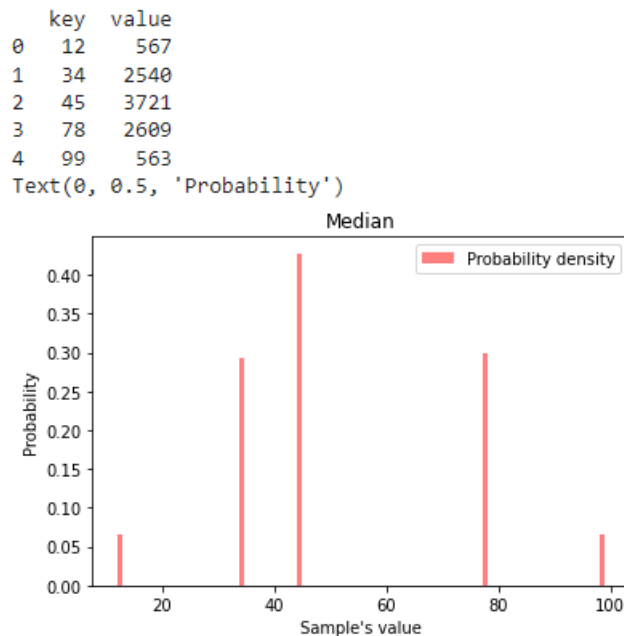
```

print(table)

# Draw the histogram of the samples
plt.hist(my_samples, bins=100, density=True, alpha=0.5, color='red')
# legend
plt.legend(['Probability density'])
# title
plt.title('Median')
# x-axis label
plt.xlabel("Sample's value")
# y-axis label
plt.ylabel('Probability')

```

Run program & result:



We will try running a sample of size 30 to see the distribution of the 5 numbers in a particular data set.

Source code:

```

import numpy as np
import matplotlib.pyplot as plt

data = np.random.choice(sample, size=30, replace=True)
bins = np.arange(100)-0.5

hist, edges = np.histogram(data, bins=bins)

```

```

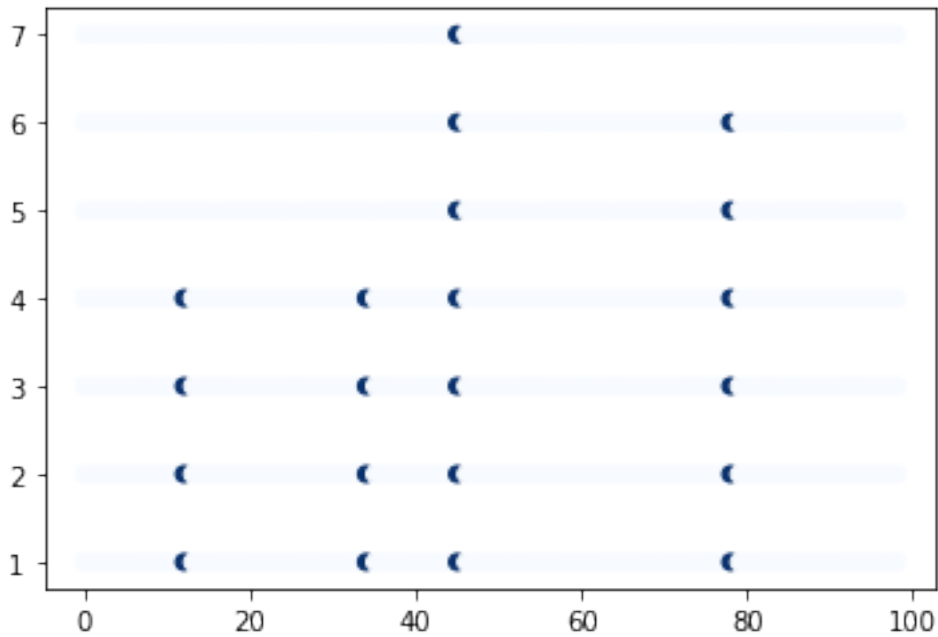
y = np.arange(1,hist.max()+1)
x = np.arange(99)
X,Y = np.meshgrid(x,y)

plt.scatter(X,Y, c=Y<=hist, cmap="Blues")

plt.show()

```

Run program & result:



The highest column is the median of the data set, which is 45 in this case. It means that the number 45 appears 7 times among the 30 numbers chosen in this data set.