

Data Science Masters :Assignment 15

Problem Statement 1:

You survey households in your area to find the average rent they are paying. Find the standard deviation from the following data: \$1550, \$1700, \$900, \$850, \$1000, \$950.

```
In [26]: # Solution
import math
import numpy as np
dataList = np.array([1550,1700,900,850,1000,950])
# Computing Avg rent paid
cumsum = 0
for num in dataList:
    cumsum = cumsum + num
    avgRent = cumsum / len(dataList)
print("Average Rent =",avgRent)
# Computing Standard Deviation
intersum = 0
for num in dataList:
    intersum = math.pow((num - avgRent),2) + intersum
variance = intersum / len(dataList)
std = math.pow(variance,0.5)
print("Standard deviation =",std)
print("-----")
# Using Numpy std. functions to cross check..
print("Using Numpy std. functions to cross check....")
print("mean=",np.mean(dataList))
print("standard deviation=",np.std(dataList))
```

```
Average Rent = 1158.3333333333333
Standard deviation = 335.92740617910624
-----
Using Numpy std. functions to cross check....
mean= 1158.3333333333333
standard deviation= 335.92740617910624
```

Problem Statement 2:

Find the variance for the following set of data representing trees in California (heights in feet):

3, 21, 98, 203, 17, 9

```
In [27]: # Solution
inputData = np.array([3, 21, 98, 203, 17, 9])
intersum = 0
cumsum = 0
for num in inputData:
    cumsum = cumsum + num
    avgHeight = cumsum / len(inputData)
for num in inputData:
    intersum = math.pow((num - avgHeight),2) + intersum
variance = intersum / len(inputData)
print("variance =",variance)
print("Using Numpy std. functions to cross check....")
print(np.var(inputData))
```

```
variance = 5183.25
Using Numpy std. functions to cross check....
5183.25
```

Problem Statement 3:

In a class on 100 students, 80 students passed in all subjects, 10 failed in one subject, 7 failed in two subjects and 3 failed in three subjects. Find the probability distribution of the variable for number of subjects a student from the given class has failed in.

```
In [90]: # Solution
totalStudents = 100
pass_allSub = 80
fail_oneSub = 10
fail_twoSub = 7
fail_threeSub = 3
probFail_zeroSub = pass_allSub / totalStudents
probFail_oneSub = fail_oneSub / totalStudents
probFail_twoSub = fail_twoSub / totalStudents
probFail_threeSub = fail_threeSub / totalStudents
print("Probability of Students failed in Zero Subject =",probFail_zeroSub)
print("Probability of Students failed in One Subject =",probFail_oneSub)
print("Probability of Students failed in Two Subjects =",probFail_twoSub)
print("Probability of Students failed in Three Subjects =",probFail_threeSub)
import matplotlib.pyplot as plt
objects = ('P(zeroFail)', 'P(oneFail)', 'P(twoFail)', 'P(threeFail)')
y_pos = np.arange(len(objects))
y=[probFail_zeroSub,probFail_oneSub,probFail_twoSub,probFail_threeSub]
plt.bar(y_pos, y, align='center')
plt.xticks(y_pos, objects)
plt.ylabel('Probability Values')
plt.title('Probability Distribution')
plt.show()
```

Probability of Students failed in Zero Subject = 0.8

Probability of Students failed in One Subject = 0.1

Probability of Students failed in Two Subjects = 0.07

Probability of Students failed in Three Subjects = 0.03

