

STATISTICS - 1 - Assignment

By Prakash Ghosh

```
In [32]: # import the libraries
import statistics as stat
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

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 [7]: # Input Data from household Survey
lst_rent =[1550, 1700, 900, 850, 1000, 950]

# standard deviation from lst_rent
result=stat.stdev(lst_rent)

print('Standard deviation for Rent is :\t', result)
```

Standard deviation for Rent is : 367.9900360969936

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 [49]: # Input Data for heights of trees in California
lst_tree_height =[3, 21, 98, 203, 17, 9]

# variance from lst_tree_height
result= stat.variance(lst_tree_height)

print('Variance for the trees in California:\t', result)
```

Variance for the trees in California: 6219.9

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 [50]: # Input Data
total_no_of_student = 100
passed_all_subject = 80
failed_1_subject = 10
failed_2_subject = 7
failed_3_subject = 3

# Probability distribution from the input Data
p_failed_0_subject = passed_all_subject/total_no_of_student
p_failed_1_subject = failed_1_subject/total_no_of_student
p_failed_2_subject = failed_2_subject/total_no_of_student
p_failed_3_subject = failed_3_subject/total_no_of_student

# Create a Dataframe with the Probability distribution and random variable
lst_no_of_Suject_failed=[0,1,2,3]
lst_probability=[p_failed_0_subject,p_failed_1_subject,p_failed_2_subject,p_failed_3_subject]

df_probability=pd.DataFrame({'No_of_Sub_Failed':lst_no_of_Suject_failed,'Probabalities': lst_probability})

print('Probability distribution of the variable for number of subjects a student has failed in:')
print(df_probability)

# Plot the Probabalitis distributions
plt.bar(df_probability.No_of_Sub_Failed,df_probability.Probabalities,width=.3)
plt.xlabel('Number of Subject Failed')
plt.xticks(lst_no_of_Suject_failed)
plt.ylabel('Probabality')
plt.title('Probabalitis distribution Plot')
plt.show()

```

Probability distribution of the variable for number of subjects a student has failed in:

	No_of_Sub_Failed	Probabalities
0	0	0.80
1	1	0.10
2	2	0.07
3	3	0.03

