Assignment 2 Basic Stats Lvl2

Set 1

In [1]:

```
#Importing the Required Libaries
import pandas as pd
import numpy as np
from scipy import stats
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

import warnings
warnings.filterwarnings("ignore")  #--to ignore warnings
```

In [2]:

```
#Qtn1.Plot the data & Find Outliers
```

In [3]:

```
#Loading the dataset
Q1 = pd.read_csv("C:/Users/Akaash/Downloads/Lvl_2.csv")     #(Filename: Lvl_2.csv)
Q1.head()
```

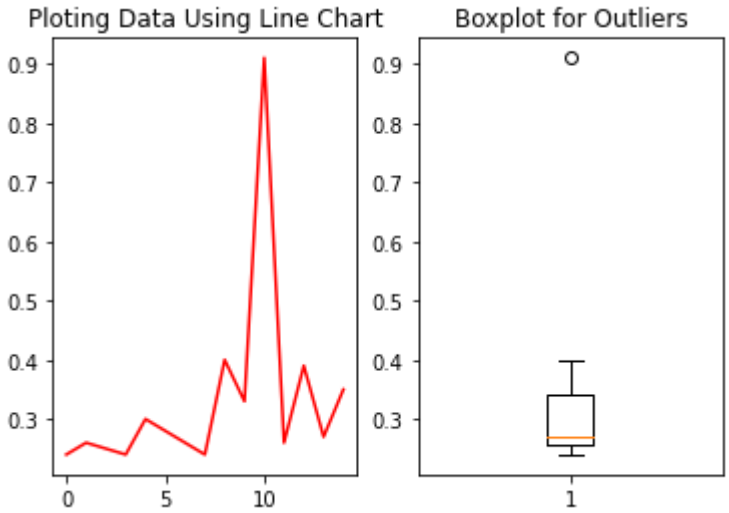
Out[3]:

	Name of company	Measure X
0	Allied Signal	0.24
1	Bankers Trust	0.26
2	General Mills	0.25
3	ITT Industries	0.24
4	J.P.Morgan & Co.	0.30

In [4]:

```
#Ploting the data - Qtn1
fig = plt.figure(figsize=(12,3))
plt.subplots(1,2)
plt.subplot(1,2,1)
plt.plot(Q1["Measure X"], color='r')
plt.title('Ploting Data Using Line Chart')
plt.subplot(1,2,2)
plt.boxplot(Q1["Measure X"])
plt.title('Boxplot for Outliers')
plt.show()
```

```
<Figure size 864x216 with 0 Axes>
```



Inference: Through line And boxplot, it can be seen there is one outlier in this dataset.

In [5]:

```
#Qtn1.find out μ,σ,σ^2
μ = Q1["Measure X"].mean()
σ = Q1["Measure X"].std()
σ2 = Q1["Measure X"].var()

print("Mean is:",μ)
print("Std is:",σ)
print("var is:",σ2)
```

```
Mean is: 0.332
Std is: 0.16853147226217766
var is: 0.028402857142857153
```

In [6]:

```
#Qtn5.iv) What is the good measure of the risk involved in a venture of this kind? Compute

#Measuring the variability of the given Probability to determine the risk involved in a ven

px = pd.Series([0.1, 0.1, 0.2, 0.2, 0.3, 0.1]) # Taking Probability of all Earning

var = px.var()

print("The Variability of the Probability is :", var)
print("The Good Measure of risk involved is:",var)
```

```
The Variability of the Probability is : 0.006666666666666666
The Good Measure of risk involved is: 0.006666666666666666
```

Set 2

In [7]:

```
#Qtn1.Given Data
μ = 45
σ = 8
#10 minutes after the car is dropped off and the customer is told that the car will be read
n = 50
```

In [8]:

```
P = stats.norm.cdf(μ,n,σ)
print("probability that the service manager cannot meet his commitment is:",P)
```

```
probability that the service manager cannot meet his commitment is: 0.265985
52904870054
```

In [9]:

```
#Qtn2.A.More employees at the processing center are older than 44 than between 38 and 44.

#P(Age>44)
A44 = 1 - stats.norm.cdf(44,38,6)
#P(38<Age<44)
A3844 = stats.norm.cdf(44,38,6) - stats.norm.cdf(38,38,6)
if A44>A3844:
    print("The Probability of having more 44 year aged people is",A44)
else:
    print("The Probability of having between 38 & 44 year aged people is",A3844)
```

```
The Probability of having between 38 & 44 year aged people is 0.341344746068
5429
```

In [10]:

```
#Qtn2.B.training program for employees under the age of 30 at the center would be expected
P36 = stats.norm.cdf(30,38,6)
print("The Probability that the employees would be under the age 30 is:",P36)
```

```
The Probability that the employees would be under the age 30 is: 0.091211219
72586788
```

In [11]:

```
#Qtn.4.Let X ~ N(100, 202). Find two values, a and b,symmetric about the mean,such that the
#taking a value between them is 0.99.
stats.norm.interval(0.99,100,20)
```

Out[11]:

```
(48.48341392902199, 151.516586070978)
```

In [12]:

```
#Qtn.5.

#Mean Profit of both Division : Mean1 + Mean2 i.e:  5 ,7
Mean = 5+7
print("Mean Profit in Rupees is",Mean*45,"Million") # Assume that $1 = Rs. 45
```

```
Mean Profit in Rupees is 540 Million
```

In [13]:

```
#SD Profit of both Division : SD1 + SD2 i.e:  3, 4
SD = 3^2 + 4^2
print("SD Profit in Rupees is",SD*45,"Million") # Assume that $1 = Rs. 45
```

```
SD Profit in Rupees is 315 Million
```

In [14]:

```
#Qtn5.A.Specify a Rupee range (centered on the mean) such that it contains 95% probability
stats.norm.interval(0.95,540,315)
```

Out[14]:

```
(-77.38865513011706, 1157.388655130117)
```

In [15]:

```
#Qtn5.C.Which of the two divisions has a larger probability of making a loss in a given yea

#Probability of making a loss for Division 1 P(x<0)
D1 = stats.norm.cdf(0,5,3)
#Probability of making a loss for Division 2 P(x<0)
D2 = stats.norm.cdf(0,7,4)
if D1>D2:
    print("The Probability of making a loss is larger for Division 1 with:",D1)
else:
    print("The Probability of making a loss is larger for Division 2 with:",D2)
```

```
The Probability of making a loss is larger for Division 1 with: 0.0477903522
728147
```

Set 4

In [16]:

```
#Qtn3.Given Data

mu = 50
s = 40
n = 100

#When X_bar = 45, for p(x<45)
x_bar = 45
t45 = (x_bar - mu) / (s/np.sqrt(100))
p45 = stats.t.cdf(t45,df=99)
print(p45)

#When X_bar = 55, for p(x<55)
x_bar = 55
t55 = (x_bar - mu) / (s/np.sqrt(100))
p55 = stats.t.cdf(t55,df=99)
print(p55)

#Calculating the probability of P(45<x<55) so,
P = 1-(p55-p45)
print(" the probability of not have an investigation is",P)
```

```
0.10712316878419327
0.8928768312158067
 the probability of not have an investigation is 0.2142463375683865
```

In [17]:

```
(45+p45)*(55+p55)
```

Out[17]:

```
2521.166879483336
```

In [ ]:

In [ ]: