**Topics: Descriptive Statistics and Probability**

1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |
|  |  |

Answer:

Measure\_X=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,

25.99,39.42,26.71,35.00])

Code:

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

#Boxplot

plt.figure(figsize=(8,4))

sns.boxplot(Measure\_X)

plt.show()

Output : Boxplot



Outlier in the Boxplot: Morgan Stanley 91.36%

Code:

Measure\_x.mean()

Mean = 33.271333

Measure\_x.std()

Standard deviation = 16.945401

Measure\_x.var()

Variance = 287.1466123809524

2.



Answer the following three questions based on the box-plot above.

1. What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

Answer:

Approximately (First Quantile Range) Q1 = 5,

(Third Quantile Range) Q3 = 12,

Median (Second Quartile Range) = 7

(Inter-Quartile Range) IQR = Q3 – Q1 = 12 – 5 = 7

And, 25 is an Outlier

1. What can we say about the skewness of this dataset?

Answer: Positively-Skewed, Median is towards the left side it is not normal distribution

1. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

Answer: In that case there will be no outliers, and it might have affected in the values of mean and median. The boxplot might have moved towards right.

3.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Answer: The mode of this data set mostly lies in between 5 to 10.

1. Comment on the skewness of the dataset.

Answer: Positively-Skewed

Mean>Median>Mode

1. Suppose that the above histogram and the box-plot in question 2 are plotted for the same dataset. Explain how these graphs complement each other in providing information about any dataset.

Answer: However, They both are Positively-skewed and both have outliers(25) we can be easily

Find the median from box plot where as in histogram mode is more visible.

4.

AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly in Polynesian. When asked about this advertisement, AT&T admitted that the portrayed incident did not actually take place but added that this was an enactment of something that “could happen.” Suppose that one in 200 long-distance telephone calls is misdirected. What is the probability that at least one in five attempted telephone calls reaches the wrong number? (Assume independence of attempts.)

**Answer:**

**Given,**

Probability of call getting misdirected (p)   = 1/200

Probability of call not getting misdirected (q) = 1-1/200 = 199/200

Number of Calls (n) = 5

n = 5

p = 1/200

q = 199/200

P(x) = At least one in five attempted telephone calls reaches the wrong number (x=1)

P(x) = (nCx) (p^x) (q^n-x)

P(1) = (5C1) (1/200)^1 (199/200)^5-1

P(1) = 0.02450

Code:

from scipy.stats import binom

Bi=binom(n=5,p=0.005)

#P(x<1)

1-Bi.cdf(0)

Output: 0.02475

5. Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

|  |  |
| --- | --- |
| x | P(x) |
| -2,000 | 0.1 |
| -1,000 | 0.1 |
| 0 | 0.2 |
| 1000 | 0.2 |
| 2000 | 0.3 |
| 3000 | 0.1 |

1. What is the most likely monetary outcome of the business venture?

Ans: The Highest probability is 0.3 for $2000

So,The most likely monetary outcome of the business venture is $2000

1. Is the venture likely to be successful? Explain

Ans: Yes, the probability that the venture will make more than 0 or a profit

p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8

So, there is a 80% chances in making profit

1. What is the long-term average earning of business ventures of this kind? Explain

Ans: The long-term average earning of business ventures of this kind is ,

E(X)=Sum (X \* P(X)) = 800$

which means on an average the returns will be + 800$

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure

Ans: The good measure of the risk involved in a venture of this kind depends on the Variability in the distribution. Higher Variance means more chances of risk

Var (X) = E(X^2) –(E(X))^2

Var(X)= (0.1⋅(−2000)^2)+(0.1⋅(−1000)^2)+(0.2⋅(0)^2)+(0.2⋅(1000)^2)+(0.3⋅(2000)^2)+

(0.1⋅(3000)^2)−(800)^2

= 400000+100000+0+200000+1200000+900000−640000

Variance = 1960000

Standard deviation(σ)=sqrt(1960000)​≈1400