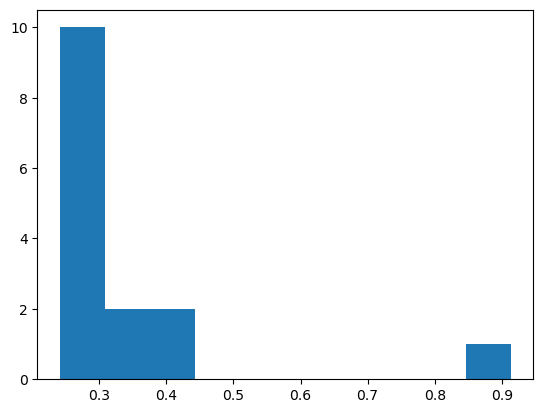
**Topics: Descriptive Statistics and Probability**

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% |



A graph of a function

Description automatically generated with medium confidence

A diagram of a chart

Description automatically generated

1. There is one outlier present in the upper extreme, which is Morgan Stanley

The mean of the data in python is given as df.mean() and mean is 26.71

1. For the standard deviation import stats module form SciPy liabrary and it is given as np.sqrt(statistics.variance(df['Measure X'])) .The standard deviation for given data is 16.94
2. The variance for given data is 287.94 by using following data as statistics.variance(df['Measure X'])



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.
2. What can we say about the skewness of this dataset?
3. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

Ans:

1. The interquartile range for the given dataset is 12-5 = 7. This value tells the range between the 25th percentile and 75th percentile value
2. The data is skewed on the right side or lower extreme since there outlier is present
3. Then the dataset become nearly normally distributed with no outliers



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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.

Ans:

1. The mode of the data is approximately at value 4-8.
2. The dataset is positively skewed since the outlier on left side.
3. From both graphs its clear that the data is positively skewed and most of the data points lies between 2 to 20.
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.)

Ans:

one in 200 long-distance telephone calls is misdirected

probability of call misdirecting  p = 1/200

Probability of call not Misdirecting = 1 - 1/200 = 199/200

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

at least one in five attempted telephone calls reaches the wrong number

= 1  -  none of the call reaches the wrong number

= 1  - P(0)

= 1   -  ⁵C₀(1/200)⁰(199/200)⁵⁻⁰

= 1  -  (199/200)⁵

= 0.02475

**probability that at least one in five attempted telephone calls reaches the wrong number = 0.02475**

1. 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?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

Ans:

E(X) = ∑X . P(X)

E(X²) = ∑X² . P(X)

Var (X) = E(X²)  - { E(X) }²

SD = √Var

|  |  |  |  |
| --- | --- | --- | --- |
| 0 | P(X) | E(X)= X . P(X) | E(X²) = X² . P(X) |
| -2000 | 0.1 | -200 | 400000 |
| -1000 | 0.1 | -100 | 100000 |
| 0 | 0.2 | 0 | 0 |
| 1000 | 0.2 | 100 | 200000 |
| 2000 | 0.3 | 600 | 1200000 |
| 3000 | 0.1 | 300 | 900000 |
| **Total** |  | 800 | 2800000 |

(i)  most likely monetary outcome of the business venture is $ 2000 as it has maximum Probability  0.3

(ii) Venture is successful if X is + ve

Hence if X is 1000 , 2000 or 3000

Probability is  0.2 + 0.3 + 0.1 = 0.6

as 0.6 > 0.5 Hence venture likely to be successful

(iii) The long-term average earning of business ventures  = E(X)

E(X) = ∑ X.P(X)  = **$ 800**

(iv)   long-term average earning of business ventures  = E(X)

E(X) = ∑ X.P(X)  = $ 800

(iv) Risk involved in a venture

Var (X) = E(X²)  - { E(X) }²

=   2800000 -   800²

= 2160000  ( Quite High)

SD = √Var  ≈ $ 1470

As Variability is Quite high  hence Risk is high