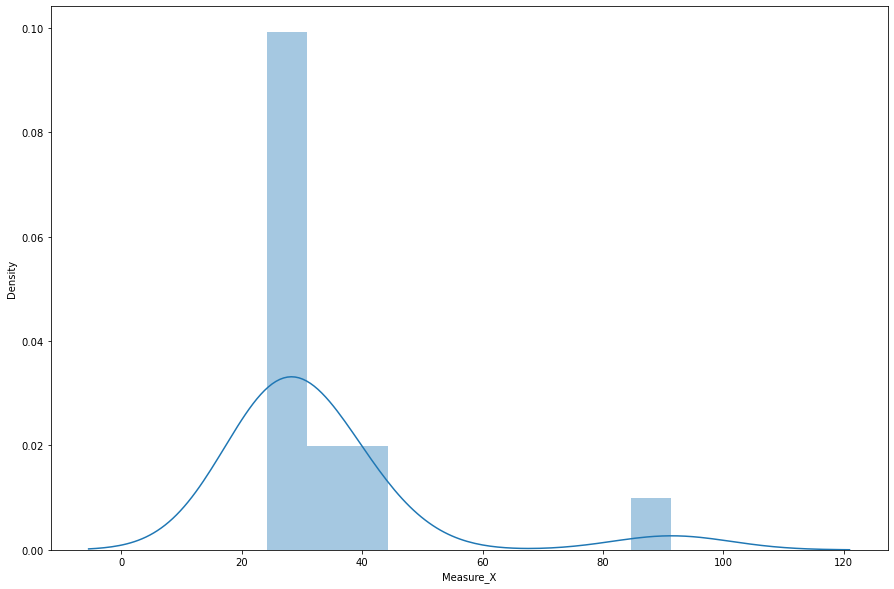
**Topics: Descriptive Statistics and Probability**

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% |
|  |  |





*First Quantile= 25.47 Second Quantile= 26.71 Third Quantile= 33.975 Inter-Quartile Range= 8.505000000000003 Upper Whisker= 46.7325 Lower Whisker= 12.712499999999995*

*The outlier1 in the boxplot: Series([], Name: Measure\_X, dtype: float64)*

*The outlier in the boxplot: 10 91.36*

*Name: Measure\_X, dtype: float64*

*count 15.000000*

*mean 33.271333*

*std 16.945401*

*min 24.140000*

*25% 25.470000*

*50% 26.710000*

*75% 33.975000*

*max 91.360000*

*Name: Measure\_X, dtype: float64*

*Variance= 287.1466123809524*



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.

*Ans: 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*

*Second Quartile Range is the Median Value*

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

*Ans: Right-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?

*Ans: In that case there would be no Outliers on the given dataset because of the outlier the data had positive skewness it will reduce and the data will normal distributed*



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

*Ans: The mode of this data set lie in between 5 to 10 and approximately between 4 to 8 .*

1. Comment on the skewness of the dataset.

*Ans: Right-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.

*Ans: They both are right-skewed and both have outliers the median can be easily visualized in box plot where as in histogram mode is more visible.*

1. 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:  IF*** *1 in 200 long-distance telephone calls are getting misdirected.*

*probability of call misdirecting   = 1/200*

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

***The*** *probability for at least one in five attempted telephone calls reaches the wrong number*

*Number of Calls = 5*

*n = 5*

*p = 1/200*

*q = 199/200*

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

*P(x) = ⁿCₓ pˣ qⁿ⁻ˣ*

*P(x) = (nCx) (p^x) (q^n-x) ###### nCr = n! / r! \* (n - r)!*

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

*P(1) = 0.0245037 (.025 after rounding)*

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 |

E(X) =Sum X.\*P(X) | E(X^2) =X^2\*P(X)

-200             | 400000

-100                 | 100000

0             | 0

200       | 200000

600         | 1200000

300         | 900000

Total: 800         | 2800000

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

*Ans: The most likely monetary outcome of the business venture is 2000$*

*As for 2000$ the probability is 0.3 which is maximum as compared to others*

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 this states that there is a good 80% chances for this venture to be making a profit*

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

*Ans: The long-term average is* ***Expected*** *value = 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:*

| **X** | **P(x)** | **E(x)** | **(x−μ)^2\*P(x)** | **Sigma^2** |
| --- | --- | --- | --- | --- |
| -2,000 | 0.1 | -200 | (-2000-800)^2\*0.1 | 784000 |
| -1,000 | 0.1 | -100 | (-1000-800)^2\*0.1 | 324000 |
| 0 | 0.2 | 0 | (0-800)^2\*0.2 | 128000 |
| 1000 | 0.2 | 200 | (1000-800)^2\*0.2 | 8000 |
| 2000 | 0.3 | 600 | (2000-800)^2\*0.3 | 432000 |
| 3000 | 0.1 | 300 | (3000-800)^2\*0.1 | 484000 |

*Mean of the data from above is 800,*

*Variance for discrete random variable is Sigma^2= SUM((x−μ)^2\*P(x)*

*Variance= sum(Sigma^2) = 2160000*

*St deviation = sqrt(Variance)= 1469.69 ( rounded to 1470)*

***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***