**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:-

Mean = 0.332

Standard Deviation = 0.1685315

Variance = 0.02840286

Outlier = 0.91



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:-The inter quartile range of the given dataset is approximately from 5 to 12, this implies that the 50% records of the datasets lies in the range of 5-12.

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

Answer:- The skewness of the dataset can also be predicted upto certain level by using this boxplot, i.e. we can say that it is positively skewed data with the long tail at the right hand side.

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 the above dataset the data point with value 25 is an outlier which mostly doesn’t affect the data, but if this value is found to be 2.5 then the inter quartile range of the boxplot might get shifted towards the RHS. It can affect the skewness as well as the centrality of the data.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans:- Mode of the data is lies in the range 4-8 of the value of Y with the frequency value of approximately being 21.

1. Comment on the skewness of the dataset.

Ans:-The skewness value of the data is negative by observing the histogram, hence it is negatively skewed data with the mass of the data being concentrated on the LHS.

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:-Histogram of the data mostly helps us to find the skewness and kurtosis of the given data.

It also gives us the most approximate values of the given dataset at any desirable point on the graph. It helps to see the peakedness of the data.

While the boxplot of the data mostly helps us to realize the concentration of the mass of the data i.e. where is the range of most of the values are. It also helps to find the IQR of the data. We can also find out if there are any outliers are present in the given data, because the outliers can cause the measure errors in our calculations while analyzing the data.

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:-

By using poissons distribution formula we can find the required probability

Here Lambda=1/40 i.e.0.025 & x varies from 1-5 hence using the formula we get following answer

=0.025\*exp(1)^-0.025+(0.025^2\*exp(1)^-0.025)/2+(0.025^3\*exp(1)^-0.025)/(3\*2)+(0.025^4\*exp(1)^-0.025)/(4\*3\*2)+(0.025^5\*exp(1)^-0.025)/(5\*4\*3\*2)

=0.02469009

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?

Ans:- Most likely monetary outcome=-2000\*0.1-1000\*0.1+1000\*0.2+2000\*0.3+3000\*0.1

=$800

1. Is the venture likely to be successful? Explain

Ans:- The long term business venture is most likely going to be successful because the probability of positive venture is more than 0.5 and the expected value is also positive $800.

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

Ans:-From the given data we can find the average venture of the business by calculating the mean of X.

-2000-1000+0+1000+2000+3000/6=3000/6

=$500

$500 is the average business venture and from this we can say that the business will most likely be profitable.

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

Ans:-From above data 40% probability of loss or no profit scenario. So this might be the risk the investor has to take in business ventures.