**M KISHORE ARVIND**

**Topics: Descriptive Statistics and Probability**

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

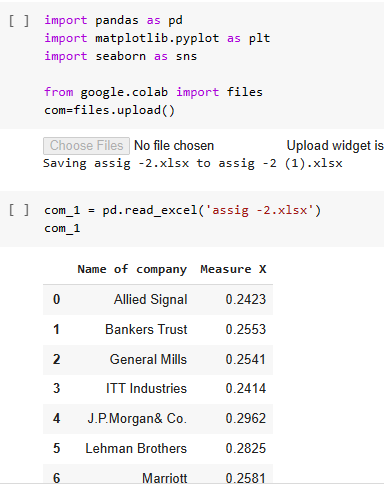
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan& Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft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

ANS):

Given is the data of a list of companies and their respective measure (X) values.

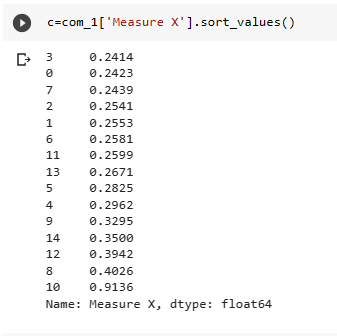
We are required to plot the data and find the mean(µ), standard deviation (σ),variance (σ^2).

Firstly, we plot the data in python notebook. Before that,we need to load this data in python by using the pandas library. Then, to plot this data points, we need to invoke the seaborn library, sns\_load\_dataset().

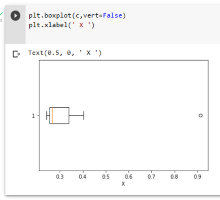




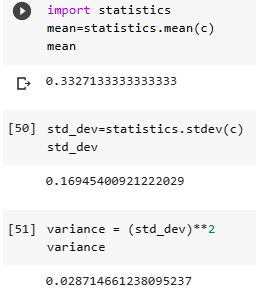
As the values in the column (X) are not in the order, we just sort the values by using sort command.



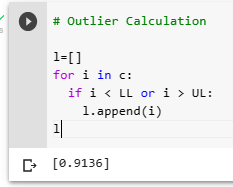
After this , we plot the this data in the box plot using the function,we invoke boxplot function from matplotlib library.



After plotting the graph, we have found out the mean(µ) , standard deviation (σ),variance (σ^2) which is given below.



To find the Outliers in the box plot .



There is only one Outlier in the Box plot which is 0.9136.

From the above screen shots, we write down the values:

Mean ( µ) = 0.3327.

Standard Deviation ( σ) = 0.169454.

Variance (σ^2) = 0.02871.

From the box plot , we get

Outlier = 0.9136.

Which means that a value of 0.9136 exist in the table which lies outside of the box plot.

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.

ANS): The inter-Quartile range of this dataset(IQR) = Q3-Q1

IQR approximately is given as = 12-5 =7 (approx).

The IQR of 7 suggest that 25% - 75% of the data lies within this region.

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

ANS): As the median line is towards the lower whisker, it is a right skewed data.

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): If the value is actually 2.5, then the point would lie inside the box plot and in the first quartile of the box plot and there would be no outliers in the box plot.All the points would lie inside the boxplot.

3. 

Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

ANS): As we can see from histogram that the highest frequency lies in the approximately 4 to 8.

1. Comment on the skewness of the dataset.

ANS): The skewness of the dataset is right skewed data.

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): Both the histogram and the box plot gives the same type of graph.

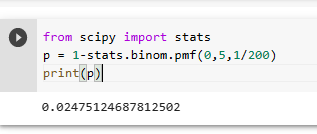
4.AT&T was running commercials in 1990 aimed at luring back customers who had switched to one of the other long-distance phone service providers. One such commercial shows a businessman trying to reach Phoenix and mistakenly getting Fiji, where a half-naked native on a beach responds incomprehensibly 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): Given to us is that 1 in 200 calls are misdirected, i.e p = 1/200 = 0.005.

No.of calls ( n ) = 5 calls

We need to find the probability of atleast one in five calls. So , the formula for atleast one call is given below.

P(X>=1).



In google colab, we import stats from scipy library and use stats.binom.pmf function to find the probability of ‘0’ calls misdirected , and minus ‘ 1’ from it.

Therefore, the Probability comes out to be (p) = 0.02475.

If we multiply with 100, we get 2.475 % chance of at least 1 in 5 calls being misdirected and reaches the wrong number.

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) : The most likely out come of the business venture is given by the expectation value

which is given as :

E(x) = ∑ (i=n) x(i)P(x(i))

Substituting the values in the equation, we get

E(x) = (-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000\*0.3)+(3000\*0.1)

E(x) = -200 + -100 + 0 + 200 + 600 + 300

E(x) = 800.

Therefore, the most likely outcome of the business venture is $800.

1. Is the venture likely to be successful? Explain

ANS) : As the expected value has come out to be 800 which is less than 1000, It is not

likely to be a successful venture .

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

Ans): Long – term average is the ratio of the Expected value and the no of samples.

Here, in this sample ,

We have no.of samples (n) = 6

The long term average earning of the business = E(x) / n

= 800 / 6 = 133.333.

Therefore, the long term average earning of the business = $ 133.33

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

Ans): A good measure of risk is always or should be maintained less than 5%.