**Topics: Descriptive Statistics and Probability**

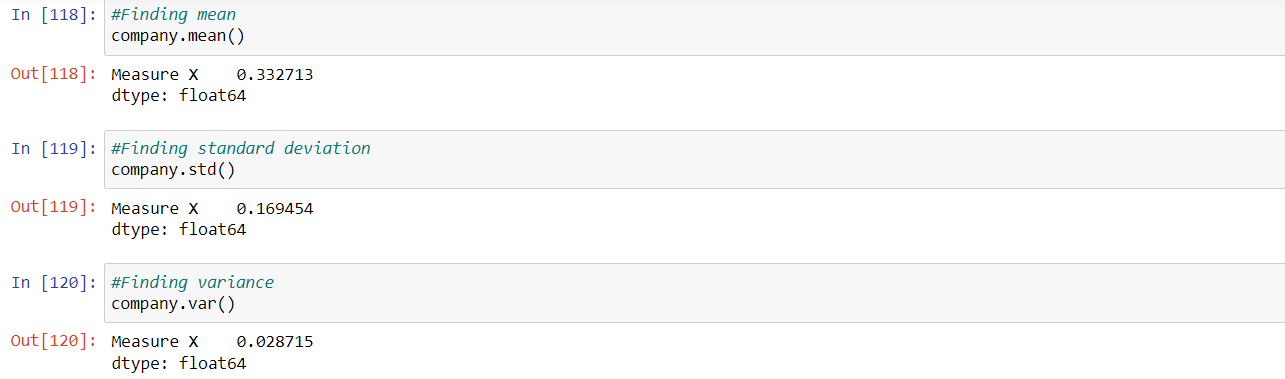
1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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% |

# Copy the table to the excel file and read the data from excel file as follows:

company = pd.read\_excel('C:\\Users\\User\\Documents\\ExcelR\\Assignments\\Assignment 2-Basic Statistics Level - 2\\company.xlsx')

**1] Finding**



**2] Finding Outliers**

****

****

**From the above observations we can say that the Morgan Stanley is an outlier of 91.36%.**



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.

**A]** IQR = 12-5=7

This value implies that 50% of the data lies within the given inter-quartile range i.e. (5 – 12).

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

**A]** We can say that the data is positively skewed. we can say that the data is more towards the right side since the median is closer to the bottom of the box plot

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?

**A]** No outliers will be present as the changed data point lies within the inter-quartile range. However the mean value would be changed. With 2.5 the lowest value, the IQR will be 12 - 2.5 = 9.5



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**A]**

|  |  |
| --- | --- |
| Class Interval ‘Y’ | Frequency |
| 0 - 2 | 2 |
| 2 - 4 | 9 |
| 4 - 6 | 21 |
| 6 - 8 | 21 |
| 8 - 10 | 15 |
| 10 - 12 | 10 |
| 12 – 14 | 8 |
| 14 – 16 | 9 |
| 16 - 18 | 4 |
| 18 - 20 | 5 |
| 24 - 26 | 1 |

From the above table we can say that most of the data lies in the interval of 4-6 and 6-8.

1. Comment on the skewness of the dataset.

**A]** The dataset is positively skewed

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.

**A]** The dataset from both the graphs are right skewed since the most of the data lies

near the median

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.)

**A]**  P (That one in 200 long distance telephone calls is misdirected) = 1/200 = 0.005

P (That one in 200 long distance telephone calls is directed) = 1-0.005 = 0.995

P (That at least one in five attempted telephone calls reaches the wrong number)

=1 – 0.9955

= 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?

**A]** The highest probability occurrence is p(x)=0.3 where X=2000

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

1. Is the venture likely to be successful? Explain

**A]** Considering the probability of

* + 1. Positive returns to be (0.2+0.3+0.1=0.6) i.e. 60%
    2. Negative returns to be (0.1+0.1=0.2) i.e. 20%.

We can say that the venture is likely to be successful.

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

**A]** E(X) = x\*P(x)

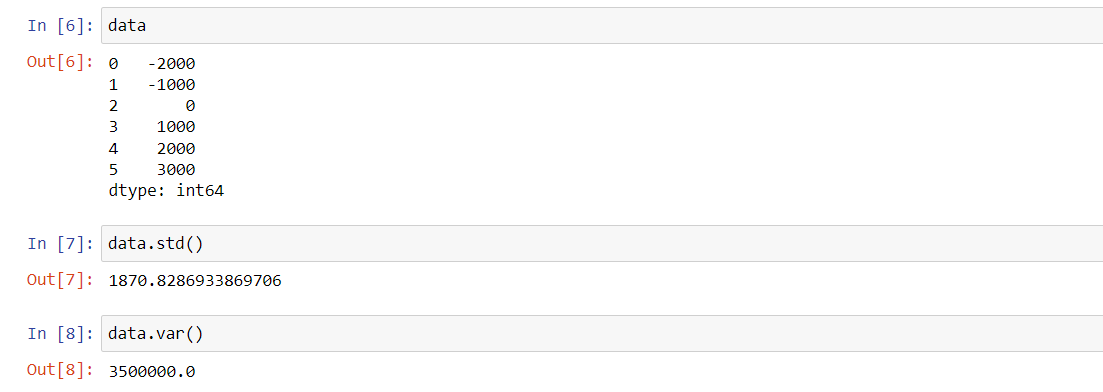
= (-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000\*0.3)+(3000\*0.1)

= 1200

The long-term average earning of business ventures of this kind is $1200

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

**A]** In order to measure the risk involved in a venture we can use standard deviation and variance

****