**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% |

**SOL:-**

**Given dataset copied and saved as Book.csv**

**1). Mean :- book['Measure X'].mean()**

**Mean = 33.2713**

**2). Standard Deviation:- book['Measure X'].std()**

**Std = 16.9454**

**3). Variance :- book['Measure X'].var()**

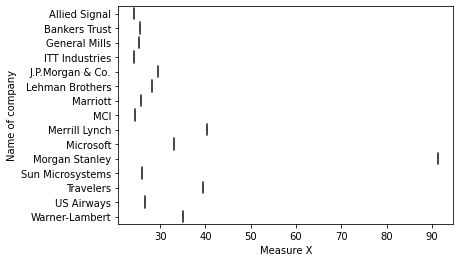
**287.1466**

**4). For detecting outlier boxplot was plotted**

**sns.boxplot(x='Measure X' , y= 'Name of company',data=book)**

**from boxplot observation**

**Morgan Stanley with 96.36% is an outlier**





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.

**Sol:- IQR = 12-5 = 7, this represents the range which contains 50% of the data points.**

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

**Sol:- Plot is right skewed**

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?

**Sol:- 2.5 will be not considered an outlier. The boxplot will start from 0 and end at 20 in representation.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Sol:- Between 4 and 8**

1. Comment on the skewness of the dataset.

**Sol:- Dataset is right 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.

**Sol:- Both Plots give idea about skewness of the data but,**

1. **By using histogram we can find frequency distribution of data points easily so we can see how many times each data point is occurring, but we cannot detect outliers with histogram We can only guess looking at the gap that 25 may be an outlier.**
2. **By using boxplot we can find outlier value easily.**
3. 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.)

**SOL:-**

**A: The call is misdirected**

**Then probability of the event A is**

**P(A)= 1/200**

**Therefore,**

**Probability that at least one in 5 attempted call reaches the wrong number**

**= 1 - Probability that no attempted call reaches the wrong number**

**= 1 – P(A bar)**

**= 1 – (199/200)\* (199/200)\* (199/200)\* (199/200)\* (199/200)**

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

**= 0.025**

**Probability that at least one in 5 attempted call reaches the wrong number = 0.025**

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?

**SOL:-**

**Max. P = 0.3 for P(2000). So most likely outcome is 2000**

1. Is the venture likely to be successful? Explain

**SOL:-**

**P(x>0) = 0.6, implies there is a 60% chance that the venture would yield profits or greater than expected returns. P(Incurring losses) is only 0.2. So the venture is likely to be successful.**

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

**SOL:-**

**Long term returns = ((-2000\*1)+ (-1000\*1)+ (1000\*2)+ (2000\*3)+ (3000\*1) / 6) = 8000/6 = 1333**

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

**SOL:-**

**Good measure is, Positive returns (profits) probability tends to be more than negative returns (loss). i.e. 60% probability of profits and 20% probability of loss.**