**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:**

Chart, box and whisker chart

Description automatically generated

Mean ( = 33.271333 %

sd ()= 16.945401 %

Var ( = 287.146612 %

Note: Jupyter Notebook file attached



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.
2. What can we say about the skewness of this dataset?
3. 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:**

1. IQR is the range between upper quartile (Q3) and lower quartile (Q1)

IQR= Q3-Q1= 12.5-5 = 7.5

Interquartile range is 7.5.

50% of data points lie between this interquartile range 7.5( that is 5 to 12.5).

ii) The distribution in positively skewed ( right skewed) .

iii) New Box plot will not have the outlier. Also Mean and Median needs to be calculated to see if there is any shift in data



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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:**

1. Mode can lie between 3 and 8 (with repeated frequency 22) because majority & repeated data lie in this range. To find out actual mode , we need to know and analyze the complete data.
2. The distribution is positively skewed (right skewed) .
3. The distribution of both is positively skewed (right skewed) . Majority of the data is distributed & concentrated towards left side with long tail on right side in both cases. Both graphs show one outlier at same location 25. All these three points complement each other & this way both graphs -Histogram and Box plot of same data complement each other. Thus Histogram and box plot indicate the same information on distribution properties of the dataset.
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:**

B = probability of 1 call misdirected in 200 long distance telephone calls

Probability of occurring of B = 1/200

P(B)= 1/200

Probability of having at least one successful call will be

1-P(B)= 1-1/200= 199/200= 0.995

As each event is independent of other event, probability that at least one in five attempted telephone calls reaches the wrong number will be

1- (0.995)^5

= 0.02475

Probability %=2.475%

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?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

**Ans:**

1. Most likely outcome will be $2000 as it has highest probability
2. Yes. Because total probability of Zero or negative Return of Venture is equal to 0.1+0.1+0.2= 0.4 while Probability of positive return of 1000 to 3000 is equal to 0.2+0.3+0.1= 0.6. There is 60% probability that the venture will be successful
3. average earning is

(-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000 \*0.3)+(3000\*0.1)=800

Long term average of this Venture will be $800.

1. Measure to evaluate the risk would be to calculate variance and standard deviation of the variable

Variance = 3500000

Standard Deviation = 1870.829

High value of standard deviation of $1870 and average returns of $800 indicates that this venture is highly risky