**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% |

**Solution:**

= Mean = 33.271

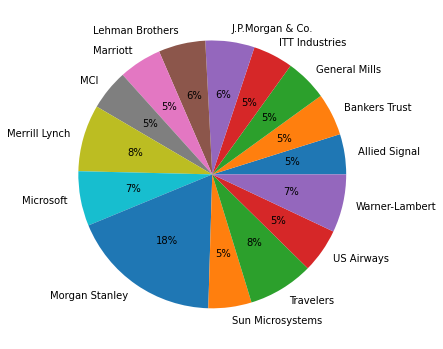
= Standard deviation = 16.370

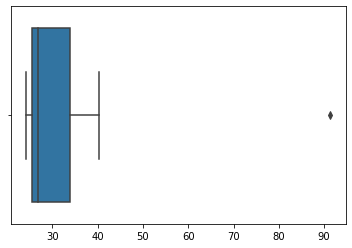
= Variance = 268.004

**Code**: Asssign2\_Set 1\_Q1 (Attached in MAIL)

The following is the outlier in the boxplot: Morgan Stanley 91.36%

The Box plot and Pie chart are presented below





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.
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?

**Solution:**

In the given dataset set :

1) To find the inter-quartile

The formula : Q3-Q1

According to the given data set Q3=12

Q1=5

From the inter-quartile formula=Q3-Q1=12-5

=7

2) When the median is closer to the bottom of the box, and if the whisker is shorter on the lower end of the box, then the distribution is positively skewed (skewed right). Hence the given dataset is **positively skewed.**

3)If the data point is changed to 2.5 instead of 25 then there will be no outliers for the dataset and because of the data is in positive skewness once the value is changed into 2.5 then the dataset will be normally distributed.

3.



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.

**Solution:**

1) According to the histogram graph the most repeated value is between 4 to 8 ,so the mode is approximately between 4 to 8

2) The dataset is Right skewed. Mean>Median>Mode

3) They both are right-skewed and both have outliers the median can be easily visualized in box plot where as in histogram mode is more visible.

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.)

**Solution:** If probability of getting 1 misdirected call out of 200 P=1/200

The probability of getting not even a misdirected call is q=1-1/200

q=199/200

So, to find the probability of getting 5 misdirected calls out of 200

will be ⁿCₓ pˣ qⁿ⁻ˣ

n=5

x=1

p=1/200

q=199/200

P(x)=5c1\*(1/200)^1\*(199/200)^5-1

P(x)= **0.0245037**

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

**Solution:**

1. E(X) =Sum X.\*P(X) | E(X^2) =X^2\*P(X)
2. -200             | 400000
3. -100                 | 100000
4. 0             | 0
5. 200       | 200000
6. 600         | 1200000
7. 300         | 900000
8. Total: 800         | 2800000
9. What is the most likely monetary outcome of the business venture?

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

As for 2000$ the probability is 0.3 which is maximum as compared to others

1. Is the venture likely to be successful? Explain

Ans: Yes, the probability that the venture will make more than 0 or a profit

p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 this states that there is a good 80% chances for this venture to be making a profit

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

Ans: The long-term average is Expected value = Sum (X \* P(X)) = 800$ which means on an average the returns will be + 800$

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

Ans: The good measure of the risk involved in a venture of this kind depends on the Variability in the distribution. Higher Variance means more chances of risk

Var (X) = E(X^2) –(E(X))^2

= 2800000 – 800^2

= 2160000