**Akash S** (akash\_sv01348)

**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)**

Outlier 91.5

mean 33.271333

var 287.146612

std 16.945401



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. Q1 = 5, Q3 = 12, Median Q2 = 7

Inter-Quartile Range IQR = Q3 – Q1 = 12 – 5 = 7

Second Quartile Range is the Median Value.

1. Right-Skewed, median is towards the left side.
2. There would be no Outliers, positive skewness would reduce and the data will be normal distributed.



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. The mode of this data set lie in between 5 to 10 (Approx 6).
2. Right-Skewed
3. They both are right-skewed and both have outliers.

Median can be easily visualized using box plot where as in histogram mode is more easily visualized.

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.)

**Ans)**   
Probability of call misdirecting = 1/200

Probability of call not misdirecting = 1-1/200 = 199/200

The probability for at least one in five attempted telephone calls reaches the wrong number, Number of Calls = 5 n = 5 p = 1/200 q = 199/200

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

P(x) = (nCx) (p^x) (q^n-x)

[ nCr = n! / r! \* (n - r)! ]

P(1) = (5C1) (1/200)^1 (199/200)^5-1 P(1) = **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

Ans)

1. 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. Yes, the probability that the venture will be successful is,

P = 0.2+0.2+0.3+0.1 = 0.8, which shows there is 80% chances for venture to make a profit.

1. The long-term average is expected value = Sum (X \* P(X)) = 800$
2. The good measure of the risk involved in a venture of this kind depends on the Variability in the distribution.

var(X) = E(X^2) –(E(X))^2 = 2800000 – 800^2 = 2160000

Higher the Variance higher the risk.