**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% |

Answer:

Outlier – 91.36%

Python Code - plt.boxplot(df.Measure X)

Mean – 0.33271333

Standard deviation – 0.16945401

Variance - 2.87146612



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?

Answer:

1. Q3 assumed as 12, so IQR is 12-5 = 7

IQR is the difference between upper quartile and lower quartile. Also called as width of the box in the box plot and used to find out how spread out the values are in the boxplot. It is used to find out the outliers by calculating whisker length.

1. Data has positive skewness, as we have outlier on end.
2. There wont be any outliers in the boxplot, if 25 value is replaced with 2.5.



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.

Answer:

1. At Frequency value >20. Y value would be approx. 4 and 8.
2. It has positive skewness, data has tail at the end.
3. 75% of the data lies under ‘12’ value of Y.

It is positive skewed data. Like Histogram has tail at the end and same boxplot has outlier at the end.

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.)

Answer is **0.025**

Probability that call is misdirected = 1/200

Probability that call is not misdirected = 1-1/200=199/200

Probability that atleast 1 in 5 attempted calls goes misdirected

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

=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?
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

Answer

1. X = 2000 with greater propability of 0.3
2. This venture is likely to be successful, because sum of probability of x positive values is greater than 0.05.

0.2+0.3+0.1 = 0.6 > 0.5

1. Average earning would be $800.

Avg.= Summation of (x\*P(x)) = 800

1. Standard deviation=1899.6, which means venture is highly risky.

Using standard deviation formula = Square root of (Summation(x-mean)\* (x-mean)\*)/(n-1))