**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% |

A) Mean = 33.7%

Std deviation = 0.169454

Variance = (std)^2 = 0.028714658116

Outliers = Morgan Stanley (91.36%)

How to find the Outliers? Found it by plotting the box plot in Excel

What method to use for others? Did it with the help of excel



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. IQR = 12-5 = 7

It implies that data close to the median is present in this range mostly. And 50% of the data set is present between 5 & 12.

1. What can we say about the skewness of this dataset?
2. It has positive skewness
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?
4. Its median would move towards left. Lower quartile would shift towards left. Upper quartile would not shift any.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. 6
3. Comment on the skewness of the dataset.
4. Positive skewness
5. 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.
6. \* Both have positive skewness.(Highest data set present on lower side in both graphs)

\* Value 25 is an outlier as it is not connected directly with other data set in both the graphs.

\* Median on the left side of the center of the graph(Because more no. of data set points are present on left side in the histogram, and the Box plot readily shows median is towards left side.)

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.)
2. Probability for a event happening and not happening is P(x) = ⁿCₓpˣqⁿ⁻ˣ

probability of a wrong number = p = (1/200)

Probability of not a wrong number = (199/200)

Probability of not connecting any wrong numbers

n = 5

x = 0(times when connecting a wrong number)

P(x) = 1 \* (1/200)^0 \* (199/200)\*5 = 0.975248753121875

Probability of connecting atleast one number = 1 – P(x) = 1 - 0.975248753121875 = 0.02475

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. 2000 $ returns
3. Is the venture likely to be successful? Explain
4. Overall Probability to get returns is 0.6 when compared to no returns which is 0.4. So mostly it can be successful.
5. What is the long-term average earning of business ventures of this kind? Explain
6. 800$. Multiplying all the possibilities
7. What is the good measure of the risk involved in a venture of this kind? Compute this measure

Risk stems from the possible variability in the expected returns. Therefore a good measure to evaluate the risk for a venture of this kind would be variance or standard deviation of the variable x.

Here,

sd= 1870 and var=3500000

The large value of standard deviation of $1870 is considered along with the average returns of $800 indicates that this venture is highly risky