**Topics: Descriptive Statistics and Probability**

1. Look at the data given below. Plot the data, find the outliers and find out

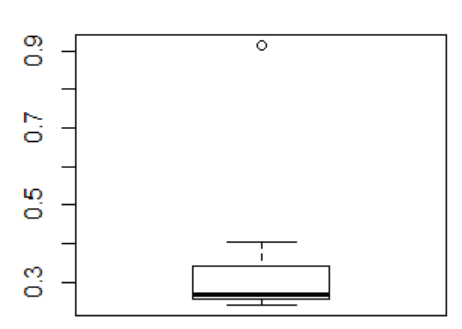
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

Ans:

Mean = 0.3327133

Variance = 0.02871466

Standard Deviation = 0.169454



So, here outlier is 0.9. Morgan Stanley is outlier.



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.

Ans: Range of IQR = (5, 12.5). It says that 50% of values lie in the range.

1. What can we say about the skewness of this dataset?

Ans: It slightly positive skewed.

1. 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: It will show a boxplot without outlier and will be more explainable.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: The highest peak of the histogram represents the location of the mode. It is between 4 and

1. Comment on the skewness of the dataset.

Ans: It is right skewed histogram with a outlier.

1. 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:

Box plot provide more of a summary of a distribution.

It typically provides the median, 25th and 75th percentile, min/max that is not an outlier and explicitly separates the points that are considered outliers.

Histogram allows to visually and quickly assess the shape of the distribution, the central tendency, the amount of variation in the data, and the presence of gaps, outliers or unusual data points.

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 at least one call reaching the wrong number is 0.025

P(E) = Probability of call being misdirected

P(Ebar) = Probability of call not being misdirected

P(Ebar) (1-P(E)) = (1-(1/200)) = (199/200)

Probability that at least one in 5 attempted call reaches the wrong number

= 1 - Probability that no attempted call reaches the wrong number

= 1-((199/200)^5)

= 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?

Ans: $2000, with the probability of 0.3

1. Is the venture likely to be successful? Explain

Ans: Yes, because while the business is having a loss of $300, it is having a profit of $1100.

So, the overall profit is of $800. So, the venture will be successful in the long term

since the probability of non-negative return is higher than 0.50.

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

Ans: The long-term average earning of business venture of this kind:

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

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

Ans:

A good measure of the risk involved in a venture of this kind is standard deviation.

Mean = 500

Std Deviation = sqrt((1/N)\*(0.1\*(-2000-500)^2)+ 0.1\*(-1000-500)^2) + 0.2\*(0-500)^2) + 0.2\*(1000-500)^2) + 0.3\*(2000-500)^2) + 0.1\*(3000-500)^2))

Std Deviation = 612