**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: - Please find the attached R file named “Ans\_Q1”.

1. mean = 0.3327133
2. Standard deviation = 0.169454
3. Variance = 0.02871466
4. Outliers = 0.9136 or 91.36%



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. IQR = 12-5 = 7 (The IQR describes the middle 50% of values when ordered from lowest to highest.)
2. Positive skewness or right-skewed data.
3. If it was found that the data point with the value 25 is actually 2.5, then the newly generated box plot will not have an outlier.



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. Mode between 4 and 8.
2. Positive skewness or right-skewed data.
3. From boxplot we can easily identify the median of the dataset while we can identify the mode of the dataset from histogram. Histogram provides the frequency of how many times each data point is occurring however boxplot provide the IQR (Inter Quartile Range) which describes the middle 50% of values when ordered from lowest to highest.
4. 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 1 call misdirected = P(X1) = 1/200

Probability of successful call out of 200 = P(X2) = 1-1/200

=199/200

= 0.995

As every event is independent of other event the probability that at least one in five attempted telephone calls reaches the wrong number will be = 1-(199/200) ^5

= 1- (0.995) ^5

= 0.0247

= 2%

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 it has the highest probability of P = 0.3
2. Yes, as the combined probability of gaining profit or return on a business venture is higher than the combined probability of gaining loss.

Probability of profit = P(Gain) = 1000 + 2000 + 3000

= 0.2 + 0.3 + 0.1

= 0.6

Probability of loss = P(Loss) = (-2000) + (-1000) + 0

= 0.1 + 0.1 + 0.2

= 0.4

Hence P(Gain) > P(Loss)

1. Long-term average earning of business = Expected Value

Expected Value = E (X) = ∑ x P (x) =

E (X) = (-2000 \* 0.1) + (-1000 \* 0.1) + (0 \* 0.2) + (1000\* 0.2) + (2000 \* 0.3) + (3000 \* 0.1)

Expected Value = E (X) = ∑ x P (x) = 800 $

Hence long-term average earning of business is $800 and the overall business venture is profitable.

1. High risk involved in a venture of this kind as it shows the high variance and negative kurtosis.

Variance = 3500000

Standard Deviation = 1870.8286

Kurtosis = -1.2