**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% |
|  |  |

* The problem is being solved in python
* Answer is as follows:

1. Outlier : Morgan Stanley - 91.36%
2. Mean : 33.2713
3. Median : 26.71
4. Standard Deviation : 16.3708
5. Variance : 268.0035



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: it is being assumed that Lowest value is 0, highest value 19, Q1 is 5, Q2 is 7, Q3 is 12.

Hence, Inter-quartile range(IQR) is:

IQR = Q3 – Q1 = 12 – 5 = 7

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

Ans: The data is skewed to the right. More data points are present towards the lower end of the range.

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: The boxplot could shift towards lower end of the data range. Q1 and Median could shift at lower part of the range.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: Mode would be between 4 to 8 value of Y.

1. Comment on the skewness of the dataset.

Ans: The data is positively skewed. It means there are more datapoints towards the lower end of the distribution.It is not symmetrical.

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: The box plot shows the total range of the data which the histogram cannot show. The box plot shows a particular boxed area where 50% or more of the data point is concentrated, but cannot show frequency for each data point. Here histogram is helpful as it can show frequencies for each datapoint. Also, Median can be located and shown in a boxplot, which is not easy in histogram.

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: let P(A) = probability of misdirection = 1/200

P(B) = probability of call reaching at correct number = 199/200

P(C) = Probability of one in five calls being correct (As all calls are independent)

= 199/200 \* 199/200 \* 199/200 \* 199/200 \* 199/200

= 0.9925

P(D) = Probability of one at-least one call getting misdirected

= 1 – 0.9925 = 0.0075

Hence, is the probability that at least one in five attempted telephone calls reaches the wrong number is 0.0075

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: The most likely outcome is of the business earning $2000 with highest probability of 0.3

1. Is the venture likely to be successful? Explain

Ans: Yes, because any business venture can continue to operate as long as its not getting into losses. The probability of this business venture getting into loss is 0.2 out of 1

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

Ans: Expected value of long-term average earning =

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

= $800

Long-term average earning of business ventures = $800

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

Ans: Mean = Long-term average earning of business ventures = $800

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Profit/loss(X) | P(x) | Deviation from mean (X – mean) | Squaring of deviation(x-mean)2 | (x – mean)2 \* P(x) |
| -2000 | 0.1 | -2800 | 7840000 | 784000 |
| -1000 | 0.1 | -2200 | 4840000 | 484000 |
| 0 | 0.2 | -800 | 640000 | 128000 |
| 1000 | 0.2 | 200 | 40000 | 8000 |
| 2000 | 0.3 | 1200 | 1440000 | 432000 |
| 3000 | 0.1 | 2200 | 4840000 | 484000 |

Sum of squared deviations:

=784000 + 484000 + 128000 + 8000 + 432000 + 484000 = 8928000

Standard Deviation: = 2955

Hence, $2955 deviation is the risk involved in a venture of this kind