**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% |

1. According to the boxplot, there’s a single outlier which is Morgan Stanley – 91.36%.

The mean of the data is 0.3327 i.e., 33.27%.

The standard deviation of the data is 0.1694 i.e., 16.94%.

The Variance of the data is 0.0287 i.e., 2.87%.



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. The inter-quartile range of the above dataset is approximately 7

i.e., Q3 – Q1 = 12 – 5 = 7.

The IQR implies the range between Q1 and Q3 which consists of 50% of the data.

1. What can we say about the skewness of this dataset?
2. Since the distance between the median and the maximum is more, the data is positively skewed.
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. If the data point is 2.5, then the new boxplot wouldn’t consist outlier and there would be changes in the values of mean, median, standard deviation and variance.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. The mode of the dataset would be 6.5 approximately.
3. Comment on the skewness of the dataset.
4. Since the mean of the data is greater than the median, it is positively skewed.
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. Histograms are used to determine the underlying distribution of the data. Boxplots are more useful when comparing multiple datasets. Both plots allow to visually assess the central tendency, the amount of variation in the data and the presence of outliers.

In this case, both the graphs are positively skewed and they have same outlier.

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. Given that one in 200 long-distance telephone calls is misdirected.

Probability of misdirected call, p = = 0.005

Probability of not misdirected call, q = 1 – p = = = 0.995

Now, the number of calls = 5

We know that, P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

Here,

n = 5

p = 0.005

q = 0.995

Now, the probability that at least one in five attempted calls reaches the wrong number

= 1 – all the calls reach the right number

= 1 – P(0)

= 1 - ⁵C₀(0.005)⁰(0.995)⁵⁻⁰

= 1 – 1(1)( 0.995)⁵

= 1 – 0.97524

= 0.02475

= 0.025 (approx.)

Therefore, the probability that at least one in five attempted telephone calls reaches the wrong number = 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. Since the probability value, P(x) = 0.3 is the maximum out of all, the most likely monetary outcome of the business venture is 2000$.
3. Is the venture likely to be successful? Explain
4. Probability of the venture to be successful;

It considers the returns from 0 as the venture will make more than 0 or a profit

= 0.2+0.2+0.3+0.1

= 0.8

This means that there is 80% chance for the venture to make profit.

Hence, the venture is likely to be successful.

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

We know the expectation formula is,

=

= 800

This means that the returns will be 800$ on an average.

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure
2. The variance is directly proportional to the risk i.e., higher variance means more chances of risk. Therefore, the good measure of the risk involved in the venture depends on the variability of the data.

Now, computing the variance

.