**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% |
|  | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

1. There is a single outlier after plotting the boxplot data, i.e., The Morgan Stanley company with 91.36%.

The Mean of the given data is (0.3327)

The Standard Deviation of the given data is (0.1694)

The Variance of the given data is (0.0287)



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 IQR of given boxplot is approximately 7. i.e., (Q3-Q1 = 12-5 = 7)

The Inter Quartile Range (IQR) consists of 50% of data. The range of Q1 & Q3 is IQR.

1. What can we say about the skewness of this dataset?
2. The given dataset is “Positively Skewed”, because the distance between median and

maximum is more than the distance between median and minimum.

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?
2. If the value 25 is actually 2.5 then there will be changes in values of a new boxplot in

Mean, Median, Variance, Standard Deviations and there will be no outliers in the

boxplot.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. The mode of the given histogram is approximately 6.5
3. Comment on the skewness of the dataset.
4. The given histogram is “Positively Skewed”, because the mean > median.
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. From the both histogram and box-plot, they are Positively Skewed. Both histogram and

box-plots allow to visually assess the central tendency, the amount of variation in the

data as well as the presence of gaps, outliers or unusual datapoints. Histograms are

preferred to determine the underlying probability distribution of a data. Boxplots are

more useful when comparing between several data sets.

From the both visualizations the outliers are same, and these are positively skewed.

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 of a wrong call (p) = 1/200 = 0.005

Probability of not a wrong call (q) = 1 – 1/200 = 1 – 0.005 = 0.995 [q = 1 – p]

Now, the total number of calls (n) = 5

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

Here, n = 5

P = 0.005

q = 0.995

Probability of getting at least one wrong call of 5 is, 1 – P(x)

= 1 - Probability of not getting a wrong call of 5

= 1 – P(x) = 1 – P(0)

= 1 - ⁵C₀ (0.005)⁰ (0.995)⁵⁻⁰

= 0.02475

~ 0.025 = 2.5%

Hence, the probability of getting at least one wrong call of 5 is 0.025 i.e., 2.5%

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. The most likely monetary outcome of the business venture from the data is 2000$ as it

has a maximum probability value i.e., P(x) = 0.3.

1. Is the venture likely to be successful? Explain
2. Yes, the venture is likely to be succeeded in future because, the probability of

Non-negative returns is 0.8 (0.2+0.2+0.3+0.1 = 0.8) i.e., 80% of profit.

1. What is the long-term average earning of business ventures of this kind? Explain
2. The long-term average earning of this business venture is = 800$. So, the

long-term average returns will be 800$ .

1. What is the good measure of the risk involved in a venture of this kind? Compute this measure
2. Variability is the good measure of the risk involved in a venture. High variance means

more chance of risk,