**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 = 0.33271**

**Standard deviation = 0.16370**

**Variance = 0.0268003**

**Outliers = [ 0.9136 ]**



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 - Inter-quartile range = 12.5 – 5 = 7.5**

**IQR contain 50% of the datapoints.**

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

**Ans - Right 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 – Nothing will be affected as new datapoint won’t change anything.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans – Dataset lie between 4 to 8**

1. Comment on the skewness of the dataset.

**Ans – Above dataset is Right Skewed.**

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 –**

**In Histogram we get central tendency as Mode with peak value where as in Boxplot we get Median value as central tendency.**

**In Boxplot with the IQR and whiskers length we can get the idea of data variability where as in Histogram the width of bins will help us to asses the data variability.**

**In Histogram we can identify outliers with isolated bar from main distribution whereas in Boxplot we can easily identify outliers.**

**Skewness can be visually seen with the asymmetry of the Histogram where in Boxplot we get a hint of asymmetry if one whisker is significantly longer than other.**

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 success - p = 1/200**

**Probability of failure - q = 1 - p = 199/200**

**Using the binomial formula**

**P(x) = () px qn-x**

**Where n is the number of trials,**

**x is the number of successful trials (misdirected calls),**

**p is the probability of success,**

**q is the probability of failure.**

**= P(0) = () ()0 ()5 = ()5**

**= 0.02475**

**The probability of at least one call being misdirected is 2.475%**

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 - Here the highest probability is 0.3 which is 2000 hence 2000 is the most likely monetary outcome.**

1. Is the venture likely to be successful? Explain

**Ans –**

**Positive outcomes = 1000, 2000 and 3000**

**Sum of probabilities = 0.2 + 0.3 + 0.1 = 0.6**

**The cumulative probability of positive outcomes is greater than 0.5 hence the venture likely to be successful.**

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

**Ans –**

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

**Average Earnings = -200 – 100 + 0 + 200 + 600 + 300**

**Average Earnings = 800**

**Long term average earning of business venture of this kind is 800/-**

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

**Ans - Calculate the mean:**

**Mean(μ) = (−2000 × 0.1) + (−1000 × 0.1) + (0 × 0.2) + (1000 × 0.2) + (2000 × 0.3) + (3000 × 0.1)**

**Mean(μ) = −200 – 100 + 0 + 200 + 600 + 300 = 800**

**Calculate the squared differences and multiply by the probabilities for each outcome: (−2000 − 800)2 × 0.1 = 784000**

**(−1000 − 800)2 × 0.1 = 32400**

**(0 − 800)2 × 0.2 = 128000**

**(1000 − 800)2 × 0.2 = 8000**

**(2000 − 800)2 × 0.3 = 432000**

**(3000−800)2 × 0.1 = 484000**

**Sum up the squared differences:**

**784000 + 32400 + 128000 + 8000 + 432000 + 484000 = 1868400**

**Calculate the square root of the sum divided by the total number of observations -**

**Standard Deviation = 558.03**

**So, the standard deviation of the distribution is approximately $558.03. This measure of risk indicates the variability or spread of the monetary outcomes around the mean.**