**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% |
| JPMorgan & 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: Morgan Stanley is the outlier**

**The mean of the dataset is 33.271333**

**The standard deviation of the dataset is 16.945400921222028**

**The variance of the dataset is 287.1466123809524**



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: The inter-quartile range of this dataset is 5 to 12**

**IQR = Q3 – Q1**

**= 12 -5**

**= 7**

**The viscous is from 0 to 19, and there is one outlier present.**

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

**Ans: The skewness of this dataset is Rightly-Skewed.**

1. If it was found that the data point with the value 25 is 2.5, how would the new box-plot be affected?

**Ans: When the value 25 is 2.5, there will be no outlier.**

**When the value 25 is 2.5, it will be nearly to normally distributed.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans: - The mode of the dataset lies between 4 to 8.**

1. Comment on the skewness of the dataset.

**Ans: - The skewness of the dataset is Left-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: - We cannot differ mode in the box-plot but we can do that in histogram.**

**The Box-Plot is best used for detecting the outliers.**

**The Histogram is best used for calculating frequency.**

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. 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 monetary outcome of the business venture is 2000,**

**Because for that outcome the probability is high about 0.3 or 30%.**

1. Is the venture likely to be successful? Explain

**Ans: The venture is likely to be successful, because**

**P(x=1000) + P(x=2000) +P(x=3000)**

**= 0.2 + 0.3 + 0.1**

**= 0.6**

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

**Ans: For that, we must calculate the expected value first. So,**

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

**= 800**

**The long-term average earning of business ventures is 800.**

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

**Ans: If we have high variance means we have high risk and if we have low variance, we have low risk.**

**E(X) 🡪 Expected Value**

**Var(X) = E(X^2) – (E(X)) ^2**

**= (-2000)2 \* 0.1 + (-1000)2 \* 0.1 + (0)2 \* 0.2 + (1000)2 \* 0.2 + (2000)2 \* 0.3 + (3000)2 \* 0.1 - (800)2**

**= 2,800,000 – 640,000**

**= 2,160,000.**

**Recommendation: - The measure of risk involved in this venture is very high as, we can see that the variance is very high!**

**Correction: - P(Loss) = P(X=-2000) + P(X=-1000)**

**= 0.1 + 0.1**

**= 0.2**

**So, the risk associated with this venue is 20%.**