**Topics: Descriptive Statistics and Probability**

* **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=>**

µ = 33.27133

σ = 16.9454

σ2 = 287.1466

**outliers = 91.36**



**Answer the following three questions based on the box-plot above.**

* **What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.**

**sol=>**IQR = Q3-Q1 = 12 – 5 = 7 (approximately). It means 50% of data points lie in the range of 5 and 12

* **What can we say about the skewness of this dataset?**

**sol=>**:- The dataset is positively skewed. Tail is found extending towards right side of the curve.

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

**sol=>**The median value will remain same, but the interquartile range will change. Moreover there will not have any outlier.



**Answer the following three questions based on the histogram above**.

* **Where would the mode of this dataset lie?**

**sol=>**:- The mode lie in between 4 and 7 .

* **Comment on the skewness of the dataset.**

**sol=>**The dataset is positively skewed as most of the data points lie in the left hand side of the curve and tail is extended towards right side.

* **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.**

1. **sol=>**The histogram and box plot will help to understand the distribution of data sets.
2. From the histogram it can be understand, how much the datasets is skewed and how higher the peak of the curve is.
3. From the box plot, it can be helpful to identify the outliers in the datasets. Even the alignment of box in boxplot will help to understand the range in which maximum datasets lie and help us to identify the median through which we can understand the distribution of data above median and below median.
4. Thus both can help to understand the datsets.

* **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**.)

**sol=>**One wrong call out of 200

**Probability of wrong call** = 1/200 = 0.005

**Probability of not wrong call** = 1-0.005 = 0.995

Probability of at least one out of five is a wrong number

= 1- Probability of all five calls are not wrong numbers

= 1- 0.995^5

= 0.0247

**=2.5%**

* **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 |

* **What is the most likely monetary outcome of the business venture?**

**sol=>Expected probability =** ∑ x\*P(X) = $ 800

* **Is the venture likely to be successful? Explain**

**sol=>Yes** ,the value above indicates that it will be successful in the long run as expected value is positive.

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

**sol=>**Long term average earning = (-2000)+ (-1000)+0 + 1000+ 2000+ 3000/6

**= $500**

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

**sol=>**The good measure of risk is the standard deviation

**sd(x) = $1870.829**