**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% |

Mean: **33.27**

Std.Deviation:**16.94**

Variance:**287.14**

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.

ANS: **Inter-Quartile Range = Q3-Q1 = 12-5 =7(approx.).**

**It means 50% of datapoints lie in the range of 5 and 12**

**Inter Quartile Range is used to find whether is there any Outliers or not if the given data is inside the inter-quartile range, so there will not be any Outliers.**

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

ANS: **Positive Skewed**

The Tail is at the right side so it is right tail end

* 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:

**The value of Median will remain same, but Interquartile will Change**



Answer the following three questions based on the histogram above.

* Where would the mode of this dataset lie?

**Here, in this histogram there is two peak Bar which means Bimodal and Mode lies between [4-8].**

* Comment on the skewness of the dataset.?

**Positive Skewed.**

* 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. ?

**From Both the Bar graph and Box-plot we can say that Graph have same nature of Distribution, and also we can say that from histogram we can see the outliers for the box plot at 25 and it affects to be normally distributed because it pulls graph tails towards right side**

* 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.)

**Probability = (1/200)\*5=1/40=0.025**

* 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?

ANS: **2000 because it is having high probability than any other Mentioned**

* Is the venture likely to be successful? Explain

ANS: **Venture is successful if X is + ve**

**Hence if X is 1000 , 2000 or 3000**

**Probability is 0.2 + 0.3 + 0.1 = 0.6**

**as 0.6 > 0.5 Hence venture likely to be successful**

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

ANS: **Long-term average earning means need to count for every probability value,E(x)=$800**

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

ANS: **Prob(x>=0) = Prob(0)+Prob(1000)+Prob(2000)+Prob(3000)**

**= 0.2+0.2+0.3+0.1**

**= 0.8**