**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% |



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. What can we say about the skewness of this dataset?
3. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

\* **Answer.1 = Q1=5, Q3=12.5(approx), so IQR=Q3-Q1=7.5(approx).**

**This value gives the range or spread of the middle 50% of the observationsi.e.the middle 50% of observations is spread out over an interval of 7.5.**

**\* Answer.2 = The dataset is positively skewed or right skewed.**

**\* Answer.3 = If the data point 25 was actually 2.5, it would not affect the new boxplot as median is plotted in the boxplot not mean.The only difference will be there will be no outliers in the new boxplot.**



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

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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. \*

**Answer.1 = The mode of this dataset will lie approx between 4.8-8.0**

**\* Answer.2 = mode<median. So the dataset is positively skewed.**

**\* Answer.3 = boxplot and histogram gives an idea of the skewness and presence of outliers in a dataset.**

4 .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:**

**P(calls being misdirected)=1/200**

**Therefore, P(call not being misdirected)=1-1/200=199/200**

**Probability that at least one in 5 attempted call reaches the wrong number**

**=1 - Probability that no attempted call reaches the wrong number**

**=1-(199/200)^5**

**=0.025**

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. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

**\* Answer.1 = most likely monetary outcome= Value of X corresponding to highest P=0.3= 2000**

**\* Answer.2 = The success of venture can be found out by adding all positive probabilities =0.2+0.3+0.1= 0.6**

**So there is a 60% chance of success.**

**\* Answer.3 = Long term average earning= sum of expected values= 800$**

**\* Answer.4 = A good measure of risk involved is standard deviation.**

**Standard deviation= 1469.69$**