**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% |

**Outliers: Morgan Stanley 91.36%**

**Mean = 0.3327**

**Variance = 0.028**

**Standard Deviation: = 0.1694**



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.

**Inter-quartile range is the range between upper quartile (Q3) and lower quartile (Q1).**

**IQR= Q3-Q1= 12-5 = 7**

**50% of the data lies between IQR.**

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

**from the above boxplot we can say that the distribution of X is right-skewed or positively 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?

**If it was found that the data point is actually 2.5 instead of 25, the outlier in the boxplot will be removed.**

**Whether the median shifts or not depends on the size of the data.**

**It will reduce the right skewness of the data.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**We need to have actual data to get the exact value of the mode. The mode can lie between 4 and 10 because there are many values in this range but this is just an assumption. The 2 bars of the same height doesn’t indicate mode every time.**

1. Comment on the skewness of the dataset.

**It is right skewed or positive 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.

**from the above histogram and barplot we can confirm an outlier at 25 in Y value. Both the plots indicate the positive skewness of the dataset.**

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.)

**let us consider the probability of 1 call misdirected out of 200 as event A.**

**Probability of occurring of event A= 1/200**

**P(A)= 1/200**

**Probability of having at least one successful call will be**

**1-P(A)= 1-1/200= 199/200= 0.967**

**As every event is independent of other event the probability will be**

**1- (0.967)^5**

**0.02475 = 2% chance.**

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?

**Max. P = 0.3 for P(2000). So most likely outcome is 2000.**

1. Is the venture likely to be successful? Explain

**P(x>0) = 0.6, implies there is a 60% chance that the venture would yield profits or greater.**

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

**Weighted average = x\*P(x) = 900. This means the average expected earnings over a**

**long period of time would be 900.**

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

**P(loss) = P(x= -2000)+P(x=-1000)=0.2. So the risk associated with this venture is 20%.**