**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% |

**Ans) Mean                           =     33.27**

**Standard Deviation   =     16.95**

**Variance                      =      287.14**

**Morgan Stanley got a Measure x of 91.36%. This is an outlier as one data point it falls away from the actual data points.**



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?

**Ans) i) The inter-quartile range of this dataset is 12-5 = 7**

**IQR = 7 (approximately).**

**It implies that the range between the 25th and 75th percentile is 7.**

**ii) The median is closer to the bottom of the box i.e., mostly to the left side of the boxplot and the whisker is shorter on the lower end of the box, so the distribution is positively skewed (skewed right).**

**iii) If the data point with the value 25 is actually 2.5, then the new box-plot will have no outlier because the data points fall in-between lower limit and the upper limit.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie? **– > Ans) 4 to 8**
2. Comment on the skewness of the **dataset. – > Ans) Right skewed ( Positive skewness )**
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**. - > Ans) We cannot differentiate the mode in the boxplot but we can do it in the histogram. Boxplots are primarily used when comparing several distributions against each other.  
   They summarize key statistics from the data and display them in a box-and-whiskers  
   format. Whereas A histogram takes only one variable from the dataset and shows the  
   frequency of each occurrence.**
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) one in 200 long-distance telephone calls is misdirected  
=> probability of call misdirecting p = 1/200  
Probability of call not Misdirecting = 1 - 1/200 = 199/200**

**P(x) = ( ncx X p^x X q^n-x )**

**p = 1/200**

**q = 199/200**

**n = 5**

**at least one in five attempted telephone calls reaches the wrong number**

**= 1 - none of the call reaches the wrong number  
= 1 – P (0)**

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

**= 0.024**

**probability that at least one in five attempted telephone calls reaches the wrong number = 0.024.**

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?

**Ans) 2000. Highest probability of 0.3**

1. Is the venture likely to be successful? Explain

**Ans) The venture is likely to be successful, because P (x = 1,000) + P (x = 2,000) + P (x = 3,000)  
= 0.2 + 0.3 + 0.1 = 0.6 (60 % successful)**

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

**Ans)**

**- 2000 \* 0.10 – 1000 \* 0.10 + 1000 \* 0.20 + 2000 \* 0.30 + 3000 \* 0.10 = $800**

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

**Ans) If the venture can maintain for long term business, then eventually it will be successful since the probability of non-negative return is higher than 0.50 (that is more than 50% and the expected value for return is a positive number ($800) and the Standard Deviation is only 8%.**