**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% |

Mean = 33.27133

Variance = 268.0035

Standard Deviation = 16.37081

Outlier can be detected using boxplot i.e. 91.36



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.

ANS-> IQR = upper quartile – lower quartile = 12-5 = 7

IQR tells that majority of data is lying in this range.

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

ANS-> As we can see that most of the data is lying in the right side of median so it is right 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?

ANS->as 25 is an outlier if it is found that it is 2.5 then it will come before 5 so median will shift a little towards left and it will be a little more right skewed.

Also there would be no outlier.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

ANS->Mode = most frequency = (4 + 8)/2 = 6

1. Comment on the skewness of the dataset.

ANS - >As the mass of the distribution is concentrated on left so it is 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.

ANS->As we can get to know about the mode and median by knowing the boxplot and histogram …mode is 6 and median is 7(approx). it follows a normal distribution with positive skewness. So we can perform all the eda process which can be implemented in normal distribution .

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.)

ANS>> Prob. Of 1 call misdirected = 1/200

Prob. Of successful call = 1-1/200 = 0.967

Prob. that atleast one in five attempted calls is wrong = 1 – (0.967)^5

= 0.02475 = 2%

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

1. Is the venture likely to be successful? Explain

ANS>>Yes, the venture is likely to be successful as the probability of gain is 60% and of loss is only 20%.

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

ANS>> Long-term average earning can be calculated by

-2000(0.1) -1000(0.1) + 0(0.2) + 1000(0.2) + 2000(0.3) + 3000(0.1)

=800

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

ANS>> Here variance is a good measure of the risk involved in a venture of this kind.

Std. dev = 1707.825