**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% |

=> In above data set we can see that “Morgan Stanley 91.36 will be the outlier.



Since we have outlier in the above data, we can't find mean is not good idea. If we drop outlier means then mean is 33.27

So, we can find the median which is the best method.

Median: 26.71%

Standard deviation: 16.94

Variance: 4.11



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 for the above Boxplot is “5 to 12”.

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

IQR= Q3-Q1= 12-5 = 7

50% of the data lies between IQR.

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

=> The skewness associate with this data set is positively skewed that 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?

=> There would be no outlier and Skewness associated with this could be “Zero”. Also, data would be symmetrically distributed and Kurtosis is associated with this data set. If the data point found to be 2.5 instead of 25.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* Mode of this lies between “4 to 10”. The 2 bars of the same height don’t indicate mode every time.

1. Comment on the skewness of the dataset.

=> Skewness is right skewed or positively 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 bar plot we can confirm an outlier at 25 in Y value. Both the plots indicate the +ve 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.)

=> Ans:

Probability of call misdirecting P(E)= 1/200

Probability of call not misdirecting is P(E1) =199/200 = 0.967

Therefore

Probability that at least one in 5 attempted call reaches wrong number P(E2) =1-(199/200) ^5

P(E2) =0.025 = 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?

* Most likely monetary outcome of the business venture: with the highest probability of 0.3. i.e., $2000

1. Is the venture likely to be successful? Explain

* The probability distribution gives us an idea about the long-term chances of earning given values of returns (indicated by x). therefore, there is a 60% probability that the venture would be successful. (Note: 0.2+0.3+0.1=0.6=>0.6\*100=>60%).

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

* The long-term average earning of business venture of this kind= (0.1) (−2,000) + (0.1) (−1,000) + (0.2) (0) + (0.2) (1,000) + (0.3) (1,000) + (0,1) (3,000) = 800.

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

=> The good measure of the risk involved in a venture of this kind is standard deviation.

> SD (ex$x) 1870.829

> var(ex$x) = 3500000

The large value of standard deviation of $1870 is considered along with the average returns of $800 indicates that this venture is highly risky.