**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% |

Sol: The Outlier in the given dataset as we plot the boxplot is: **Morgan Stanley-91.36%**

Mean of the dataset = **33.2893** { np.mean(companies\_df.Measure) }  
 Variance of the dataset = **267.6793** { np.var(companies\_df.Measure) }

Standard Deviation of the dataset = **16.3909** { np.std(companies\_df.Measure) }



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?

Sol: (i) The Inter Quartile Range = Q3-Q1 = 12 – 5 = **7** (approximate values from the above boxplot). The interquartile range of a dataset contains the middle 50% values of a dataset.

(ii) As the median lies to the left of the boxplot, it is a **Right Skewed Boxplot.**

(iii)If the value 25 is actually 2.5 then there would be no outliers in the dataset and then it would lie between the ranges of the dataset.



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.

Sol: (i) The Mode of the above dataset lie approximately between **4 to 8.**

(ii) The dataset is **Right skewed** as Mean>Median>Mode.

(iii)Comparing the Histogram and boxplot for question 3 and 2 we can say that they both are right skewed by their appearance and have outliers. The Median is easily visible in the boxplot whereas the Mode is easily visible in the histogram.

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.)

Sol: Given: If One in 200 long-distance calls are getting misdirected.

P (At-least one in five calls getting misdirected) =?

Number of calls = n = 5

p = 1/200 (Probability of calls misdirecting)

q = 1-p = 199/200

P (at-least one in five calls misdirecting) = P (x) = P (1)

Probability formula = **nCx px qn-x**

P (1) = 5C1 \*(1/200)1 \*(199/200)5-1

**P (1) = 0.024503**

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

Sol: (i) the most likely monetary outcome of the business venture is **2000$** as the probability for 2000$ is 0.3 which is maximum compared to others.

(ii) Yes. The probability that the venture will be successful is (profit or more than 0) = p(x>0) + p(x>1000) + p(x>2000) + p(x>3000) = 0.2+0.2+0.3+0.1 = 0.8 which states that there is **80% chances of making a profit.**

(iii) Long-term average earning of business venture is Expected value = ∑(X\*P(X)) = **800$.**

(iv) Good measure of the risk involved in a venture depends on the variability of the distribution. High variance denotes high risk.

Variance = ∑(X2\*P(X))- (µ)2 = = ∑(X2\*P(X))- { ∑(X\*P(X)) }2 = 2800000 – 8002

**variance(x) = 2160000.**