**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 : For this question I have mentioned the solution in ipynb file



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= Q3-Q1

= 12-5

= **7**

* This value represents the range where the middle 50% of the data lies between 5 and 12
* A larger IQR indicates greater variability in the central position of the dataset.

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

Ans (ii) : Based on the box plot, right whisker is longer than the left one, that means the data is 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?

Ans (iii) :

* If the data point with the value 25 is actually 2.5 then the box plot will be affected significantly.
* The right whisker will be shorter than before since, the box plot is sensitive to extreme values and the value 25 had a major impact than that of 2.5
* There will be no outlier in the new dataset.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: The mode of this data set lie in between 5 to 10 and approximately between 4 to 8 .

1. Comment on the skewness of the dataset.

Ans (ii)

* The dataset is positively skewed.
* Histogram is skewed to the right and the tail extends to the right and bulk of the distribution is on the left .

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:

* Both graphs are right skewed
* Both graphs have outlier
* the median can be easily visualized in box plot where as in histogram mode is more visible.

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 3) Let P be the probability of a single call being misdirected which is = 1/200

Probability of a single call not misdirected is = 1-1/200= 199/200

The probability that none of the five calls are misdirected is = (199/200)5

Probability that at least one in five attempted telephone calls reaches

The wrong number is = 1-(199/200)5

= 1-0.9752

**= 0.02475**

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 : Most likely monetary outcome of the business venture is 2000 as it is having maximun

Probability = **0.3**

|  |  |  |
| --- | --- | --- |
| x | P(x) | E(x)P(x) |
| -2,000 | 0.1 | -200 |
| -1,000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 3000 | 0.1 | 300 |
|  |  | 800 |

1. Is the venture likely to be successful? Explain

Ans : ∑E(x)P(x)= 800

Long-term average earning of business venture= 800

Venture is likely to be successful as expected value is + 800

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

Ans : E(x)= ∑X.P(x) = **800**

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

Ans: Higher Variance means more chances of risk

Var (X) = E(X^2) –(E(X))^2

= 2800000 – 800^2

= **2160000**

Hence, there is a risk