**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% |

Answer: Refer Jupyter Notebook



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.

**Answer:**

* Q1 ( Lower quartile ) = 5
* Q2 ( Median/ 2nd quartile ) = 7
* Q3 (Upper quartile ) = 12

IQR (Inter-quartile range) = Q3 – Q1 = 12 -5 = 7

This value implies that it is equal to median/ 2nd quartile value

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

**Answer:** Right skewed which is not normal distribution

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?

**Answer:**

* In that case there would be no Outliers on the given dataset.
* The Data will have normal distribution.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Answer:** The mode of this dataset of this histogram is between range 4-10 approx

1. Comment on the skewness of the dataset.

**Answer:** The skewness of this datset is Right Skewed

**Where mean > median > mode**

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.

**Answer:**

* Both are right skewed.
* Box plot is mostly used to detect the outliers & In this both graphs outliers can be easily visible.
* Box plot provide the median value.
* Histogram mostly gives us most frequently value range with interval. So, from histogram we can identify the mode value range.

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.)

**Answer:**

* 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
* Number of Calls = 5
  + P(x) = ⁿCₓpˣqⁿ⁻ˣ
  + n = 5
  + p = 1/200
  + q = 199/200
* 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   -  ⁵C₀(1/200)⁰(199/200)⁵⁻⁰
* = 1  -  1 x 1 x (199/200)⁵
* = 0.02475
* **Probability that at least one in five attempted telephone calls reaches the wrong number = 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?

**Answer:** Most likely monetary outcome of the business venture  is $2000 as it has maximum probability = 0.3

1. Is the venture likely to be successful? Explain

**Answer:**

|  |  |  |
| --- | --- | --- |
| E(x) | P(x) | E(x)P(x) |
| -2000 | 0.1 | -200 |
| -1000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 3000 | 0.1 | 300 |

* Expected value =  ∑E(X)P(X)  = 800
* Venture is  likely to be successful as Expected value is + ve   = $ 800

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

**Answer:** Long-term average earning of business ventures  = $ 800

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

|  |  |  |  |
| --- | --- | --- | --- |
| x | P(x) | x.P(x) | X² . P(X) |
| -2000 | 0.1 | -200 | 400000 |
| -1000 | 0.1 | -100 | 100000 |
| 0 | 0.2 | 0 | 0 |
| 1000 | 0.2 | 200 | 200000 |
| 2000 | 0.3 | 600 | 1200000 |
| 3000 | 0.1 | 300 | 900000 |
| **Total** | **=** | **800** | **2800000** |

* Risk involved in a venture

Var (X) = E(X²)  - { E(X) }²

=   2800000 -   800²

= **2160000**  ( Quite High)

SD = √Var  ≈ **$ 1470**

As **Variability is Quite high**  hence **Risk is high**