**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% |

**Solution:**

1. **Manual Calculation**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Given Data Set** | | | | | | | | | | | | | | | |
| 24.23% | 25.53% | 25.41% | 24.14% | 29.62% | 28.25% | 25.81% | 24.39% | | 40.26% | 32.95% | 91.36% | 25.99% | 39.42% | 26.71% | 35.00% |
| **Arranging in Ascending Order** | | | | | | | | | | | | | | | |
| 24.14% | 24.23% | 24.39% | 25.41% | 25.53% | 25.81% | 25.99% | 26.71% | | 28.25% | 29.62% | 32.95% | 35.00% | 39.42% | 40.26% | 91.36% |
| **Finding Median or Second Quartile (Q2)** | | | | | | | | | | | | | | | |
| 24.14% | 24.23% | 24.39% | 25.41% | 25.53% | 25.81% | 25.99% | 26.71% | | 28.25% | 29.62% | 32.95% | 35.00% | 39.42% | 40.26% | 91.36% |
| Finding First Quartile (Q1) | | | | | | | Q2 | | Finding Third Quartile(Q3) | | | | | | |
| 24.14% | 24.23% | 24.39% | 25.41% | 25.53% | 25.81% | 25.99% | 26.71% | | 28.25% | 29.62% | 32.95% | 35.00% | 39.42% | 40.26% | 91.36% |
| **Calculating Maximum & Minimum** | | | | | | | | | | | | | | | |
| Minimum =max(Q1-1.5(IQR), min\_value) Minimum = max (25.41%-1.5(9.59%),24.14%) Minimum = max (11.025%,24.14%) Minimum = 24.14% | | | | | Inter Quartile Range (IQR) = 9.59% | | | | | | Maximum = min(Q3-1.5(IQR), max\_value) Maximum = min (35% + 1.5(9.59%),91.36%) Maximum =min (49.38%,96.36%) Maximum = 49.38% | | | | |
| 24.14% | 24.23% | 24.39% | 25.41% | 25.53% | 25.81% | 25.99% | 26.71% | | 28.25% | 29.62% | 32.95% | 35.00% | 39.42% | 40.26% | 91.36% |
|  | | | | | | | |  | | | | | | | |
| * Mean(µ) = = = 33.271% * Variance (2) =  = = 21.72% * Standard Deviation = =4.66% | | | | | | | | | | | | | | | |

1. **By Python**

|  |
| --- |
|  |
|  |

|  |
| --- |
|  |
|  |
|  |



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.

|  |  |
| --- | --- |
| Q1 = 5 | 1st Quartile of the data set which means 25th Percentile value or (1st part when we divide data in 4 parts |
| Q2 = 7 | 2nd Quartile of the data set which means 50th Percentile (Median) value or 1st & 2nd part when we divide data in 4 parts |
| Q3 = 12 | 3rd Quartile of the data set which means 50th Percentile value or 1st,2nd & 3rd part when we divide data in 4 parts |
| IQR = Q3 – Q! = 12 – 5 = 7 | The Range between 50th & 25th percentile |

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

**Right Skewed data or Positively Skewed data:** At Right side of the Median, the difference between each data points is more as compared to left side of the median.

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?

If that is the case,

* The Q2 (Median) will move towards the center of the box,
* Right Skewness or Positive skew will reduce.
* May be the data will become Normally Distributed



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* Mode will be Between 4 to 7.5 (approximate numbers)

1. Comment on the skewness of the dataset.

* **Right Skewed data or Positively Skewed data:** At Right side of the Median, the difference between each data points is more as compared to left side of the median.

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.

|  |  |
| --- | --- |
| Box Plot | Histogram |
| It shows Outlier by plotting dot outside of the maximum whisker | It shows small frequency with far distance from the median |
| It shows the skewness of the data by 25th percentile line being nearer to q1 | It shows the skewness of the data by frequency of the data between 0 to 22 |
| Mainly provides the information about Outliers | This provide information about whether the data is continuous or discrete, How much frequency is there in bins, shape of data |

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.)

Probability of Success = p = = = 0.005

Probability of failure = q = = = 0.995

P (Misdirected 1) = (5C1 + 5C2 + 5C3 + 5C4 + 5C5) \* pr \* qn-r

= 1 – pourability of Number of calls not misdirected

= 1- nCr \* pr \*qn-r

= 1- 5C0 \* (0.005)0\*(0.995)5

= 1 - 0.9752= 0.024 = 2.4%

**Answer: The probability that at least one in five attempted telephone calls reaches the wrong number = 0.024**

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?

* As the probability (0.3) is more for 2000 $ as compared to others,
* Therefore, most likely monetary outcome of the business venture = 2000$

1. Is the venture likely to be successful? Explain

* Long term average = = (-2000\*0.1) +(-1000\*0.1) +(0) +(1000\*0.2) +(2000\*0.3) +(3000\*0.1) = 800$
* As the long-term average gives positive numbers the Business venture likely to be successful.

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

* Long term average = = (-2000\*0.1) +(-1000\*0.1) +(0) +(1000\*0.2) +(2000\*0.3) +(3000\*0.1) = 800$
* Means on an average Return will be 800 $

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

* They will get average return of 800$