**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% |

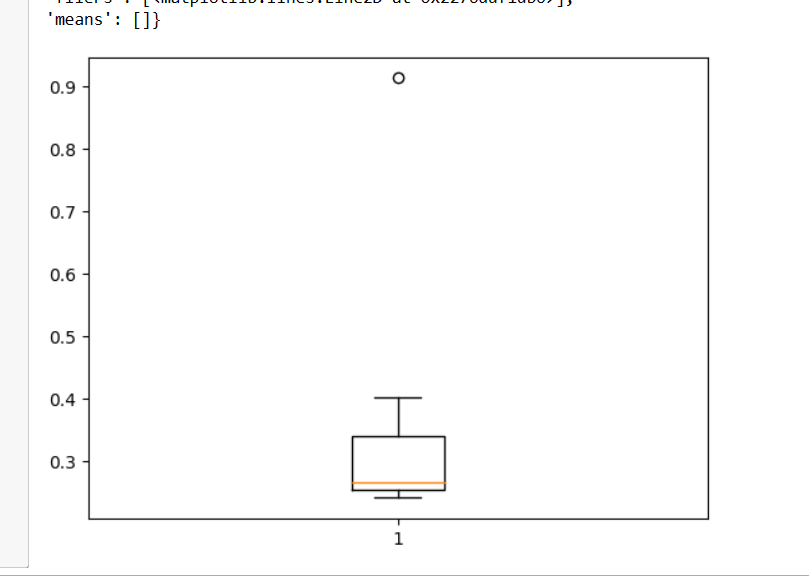
**Ans:**

The following is the outlier in the boxplot = 0.9

Mean() = 0.332

Variance(σ2) = 0.028

Standard deviation (σ) = 0.169





**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:**

Approximately (First Quartile Range) Q1 = 5

Approximately (third Quartile Range ) Q3 = 12

Median ( Second Quartile Range) = 7

Inner Quartile Range IQR = Q3 – Q1

= 12 – 5

= 7

Second Quaratile Range is the Median value.

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

**And:**

Right-skewed median is towards the left side it is not nominal 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?**

**Ans:**

In that case there would be no outliers on the given database because of the outlier the data had positive skewness it will reduce and the data will normal distributed.



**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:**

Right-skewed. 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.**

**Ans:**

Median in boxplot and mode in histogram. Histogram provides the frequency distribution so we can see how many time r=each data point is occurring however, boxplot provides the quintile distribution i.e 50%. They both are right-skewed and both have outliers 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:**

If in 200 long distance telephone calls are getting misdirection.

Probability of calls nisdirecting – 1/200

Probability of call not misdirectiong = 1 –

=199/200

The probability of at least one in five attempted telephone calls reaches the wrong number

Number of calls = 5

n = 5

p = 1/200

q = 199/200

p(x) = at least one in five attempted telephone calls reaches the wrong number

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

P(x) = (nCx) (p^x) (q^n-x) # nCr = n! / r! \* (n - r)!

P(1) = (5C1) (1/200)^1 (199/200)^5-1

P(1) = 0.0245037

The probability that at least one in five attempted telephone calls reaches the wrong number is 0.0245037

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:**

The most likely monetary outcome of the business venture is 200$. As for 2000$ the probability is 0.3 which is maxmum as compared to others.

1. **Is the venture likely to be successful? Explain**

**Ans:**

Yes, the probability that the vrnti=ure wil make more than o or a profit p(x>00)+p(x>1000)+p(x>2000)+p(x>3000)

= 0.2+0.2+0.3+0.1

= 0.8

This states that there is agood 80% changes for this venture to be making a profit

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

**Ans:**

The long-term average is Expected value = Sum (X \* P(X)) = 800$ which means on an average the returns will be + 800$.

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

**Ans:**

The good measure of the risk involved in aventure iof this kind depends on the variability in the distribution. Higher variance mrans more chances of risk.

Var(x) = E(X^2) – (E(X))^2

= 2800000 -800^2

= 2160000.