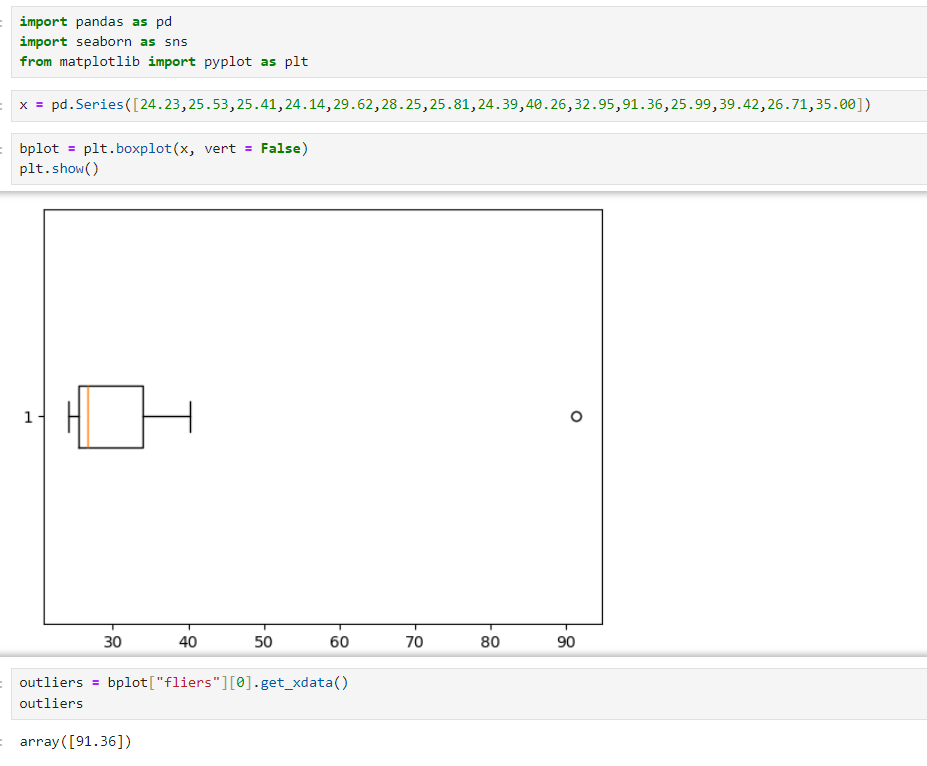
**Topics: Descriptive Statistics and Probability**

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%** |
| **JPMorgan & 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 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?**

**Answers**

**(i)** IQR = Q3-Q1

= 12-5

= 7

7 is the median of the dataset.

**(ii)** This dataset is slightly right skewed. Since, most of the data is concentrated on the left side.

**(iii)** If new data point is actually 2.5, then there will be no outlier. And the data may become normally distributed.



**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.**

**Answers**

1. The mode of this dataset would lie in range 5 to 10
2. Y values are Right Skewed. With most of the observation’s concentrated towards the left side.
3. Both the plots are right skewed. Histogram helps us in finding Mode and Barplot helps in finding Median for the dataset.
4. **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**

No. of trails = 5

Probability of Failure (q) = 1/200

Probability of Success (p) = 1-q

= 199/200

According to Binomial Distribution,

**P(x) = ⁿCₓ pˣ qⁿ⁻ˣ**

P(x>=1) = 1- p(x=0)

= 1-qⁿ

= 1- (199/200)5

= 0.0247

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**

**Answers**

1. The most likely monetary outcome of the business is $2000. Since, it has the highest probability.
2. This Business venture can be considered successful if it makes profit ≥0.

There is 80% chance that this business venture will be successful.

(0.2+0.2+0.3+0.1 = 0.8, i.e., 80%)

1. **E(x) = ∑ xi \* P(xi)**

= $800

**∴** The long term average earning = $800

Which indicates on an average the returns will be $800

1. The amount of risk involved depends on the variance of the probability distribution.

**Variance (X) = E(x2)-(E(x)) 2**

**E(x2)** = 0.1 \* (-2000)2 **+** 0.1 \* (-1000)2 **+** 0.2 \* 0 **+** 0.2 \* (1000)2 **+** 0.3 \* (2000)2 **+** 0.1 \* (3000)2

= 400,000 + 100,000 + 200,000 + 1,200,000 + 900,000

= 2,800,000

**[E(x)] 2** = 8002

= 1, 60,000

**∴ Variance(X)** = 2,800,000 - 1, 60,000

= 2,160,000