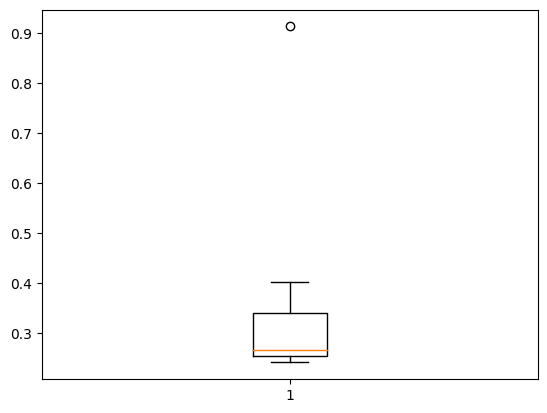
**Topics: Descriptive Statistics and Probability**

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:



* Outlier is Morgan Stanley with measure 91.36%

Using df.describe()

* Mean () = 0.332713
* Standard deviation () = 0.169454
* Varience ()= 0.0287



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?

Ans:

1. IQR = 12-5 = 7(Approximate value). It means that 50% of data points are lie between 5 and 12
2. Tail is extended right side of the curve, so it is positively skewed
3. If the value is 2.5 instead of 25, the value between 0 and 5 that is no outlier is formed and it is 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.

Ans:

1. The mode of the data set lie in between approximately 4 and 8
2. It is right skewed, Outlier is present in right side or tail extended right side (Mean>median>mode).
3. Both graphs have outliers and right skewed. Median is more visualized in boxplot but in histogram mode is more visible. Boxplot it shows the IQR , min and max values of datapoints, In histogram also visualize the frequency of each datapoints.
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.)

Ans: one in 200 long-distance telephone calls is misdirected.

Probability of call misdirecting = 1/200

Probability of call not misdirecting = 1- 1/200 = 199/200

Number of calls = 5

At least one in five attempted telephone calls reaches the wrong number P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

=5c1\*(1/200)^1\*(199/200)^4

=0.0245037

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?
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

Ans:

1. The most likely monetary outcome of the business venture is probability 0.3(p(0.3))because of high monetary value.
2. Yes, Loss will be = 0.1+0.1 = 0.2 that is 20% chance to loss. The probability that the venture will make more than 0 or a profit p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 this states that there is a good 80% chances for this venture to be making a profit
3. The long-term average is Expected value Sum(x\* P(x)) = -2000\*0.1+-1000\*0.1+0\*0.2+1000\*0.2+2000\*0.3+3000\*0.1 = 800 Which means on an average the returns will be + 800$
4. The good measure of the risk involved in a venture of this kind depends on the Variability in the distribution. Higher Variance means more chances of risk Var (X) = E(X^2) –(E(X))^2 = 2800000 – 800^2 = 2160000.(E(X²) = ∑X² . P(X))

|  |  |  |  |
| --- | --- | --- | --- |
| X | P(X) | E(X)= X . P(X) | E(X²) = 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 |