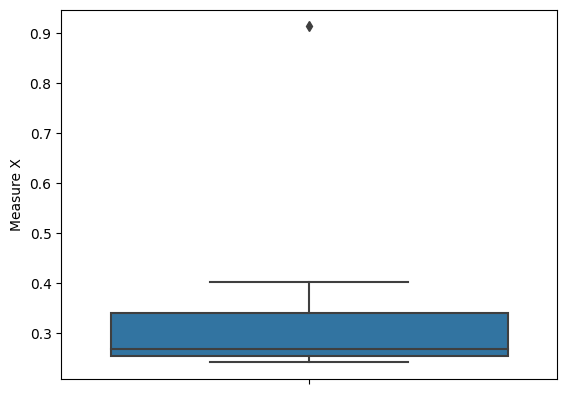
**Topics: Descriptive Statistics and Probability**

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% |

* df["Measure X"] = df["Measure X"].str.rstrip("%").astype("float") / 100
* pd.to\_numeric(df["Measure X"])
* sns.boxplot(y="Measure X", data=df)



* df["Measure X"].mean() 0.3327133333333333
* df["Measure X"].std() 0.16945400921222029
* df["Measure X"].var() 0.028714661238095233



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.

* IQR= Q3 – Q1 = 12 – 5 = 7
* IQR , It represent middle 50% data of the dataset.

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

* Here, from the boxplot ,we can say that data is right skewed
* Here, we may conclude that mean of the dataset is less than 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?

* The new Box plot will not have any outlier in the data
* It will affect skewness, The new data will be normally distributed, there will be no skewness in the dataset.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* The mode of the dataset lies between 4 and 8.

1. Comment on the skewness of the dataset.

* The data is right skewed.

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.

* 1.Box-plot provides information of outliers in the dataset .it can’t give information of central tendency of the dataset.
* 2.Histogram provides information of the frequency of the datapoints in the dataset,

It also provides information of mean, median and mode, it does not give outliers information.

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.)

* The call is misdirected then probability of the event X is

P(X)= 1/200 =0.005

* Probability that at least one in 5 attempted call reaches the wrong number

= 1 - P(E)

= 1 - (199/200)5

= 1 – 0.975234

= 0.025234

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?

* When value of x =2,000 , it has the highest probability in the dataset.

1. Is the venture likely to be successful? Explain

* There are 60% chances of being called as successful venture.
* Probability of being called successful = (0.2)for 1000 + (0.3) for 2000 + (0.1) for 3000

=60%

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

* Long term earnings = [(-2000\*0.1) + (-1000\*0.1) +(0\*0.2)+(1000\*0.2)+(2000\*0.3)+(3000\*0.1)]

= 800

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

* As, the probability of venture for being successful is 60% and for being unsuccessful is 40%.so according to probability ,there is high possibilities of being successful.