**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 :**

df["Measure X"] = df["Measure X"].str.rstrip('%').astype('float') / 100.0

df["Measure X"].agg(["mean","std","var"])

mean 0.332713

std 0.169454

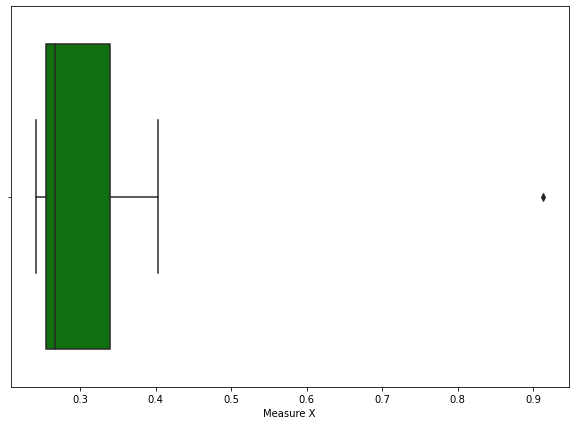
var 0.028715

Name: Measure X, dtype: float64

plt.figure(figsize =(10, 7))

sns.boxplot(x="Measure X", data=df ,color = "green",)

plt.show()



Q25per=df["Measure X"].quantile(0.25)

Q75per=df["Measure X"].quantile(0.75)

IQR = Q75per - Q25per

IQR

0.08504999999999996

Q3 = Q75per + 1.5 \* IQR

Q1 = Q25per - 1.5 \* IQR

df[(df["Measure X"] < Q1) | (df["Measure X"] > Q3)]

Name of company Measure X

10 Morgan Stanley 0.9136



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 of this dataset is 7 [Q1=5,Q3=12]. The IQR gives the range between upper

quartile and lower quartile. Data points falling outside this range are known as

outliers and approximates the amount of 50% of the data lies between IQR.

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

* Positive skewness or Right skewness

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 range of the upper quartile and lower quartile is 0 ≤ X ≥ 19 so There will be

no outlier if the value of 25 was actually 2.5.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* Mode of the dataset = 4 to 7

1. Comment on the skewness of the dataset.

* Positive skewness or Right skewness

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.

* Both plots are positively skewed, there remains an outlier of the value of 25.

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.)

* **Solution :-**
  + - * + 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 – 0.9752

= 0.025

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?
   * + The most likely monetary outcome of the business venture is 2000$ as it

has maximum Probability  0.3.

1. Is the venture likely to be successful? Explain

* Probability of success = 0.6 [0.2+0.3+0.1]

so, 60% chance that the venture would be successful.

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

* [(-2000\*0.1)+(-1000\*0.1)+(0\*0.20)+(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

* A good measure to evaluate the risk would be variance and standard deviation of

the variable x

Var = 3500000

Sd = 1870.83

The large value of standard deviation of $1870 is considered along with the average returns of $800 indicates that this venture is highly risky