**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% |

Name= ['Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan & Co.','Lehman Brothers','Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Travelers','US Airways','Warner-Lambert']

y = [24.23, 25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35]

plt.pie(y ,labels = Name, autopct='%.0f%%', explode= [0,0,0,0,0,0,0,0,0,0,0.3,0,0,0,0])

plt.show()



dk=sns.boxplot(data=df, x = 'Measure X');



The following is the outlier in the boxplot: Morgan Stanley 91.36%

y.mean()

**Mean = 33.271333**

y.std()

**Standard deviation = 16.945401**

y.var()

**Variance = 287.1466123809524**



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:** IQR = Q3-Q1 = 12- 5 = 7

Inter-quartile range of this dataset is 7. 50% datapoints lies in IQR.

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

**Ans:** Skewness of above boxplot is positive skewed as median is near to lower quartile.

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:** if 25 value is actually 2.5 then new boxplot doesn’t occur any outlier as 2.5 value lies in lower extreme nearly.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans**: Mode of this dataset lies in [4,8] approximately.

1. Comment on the skewness of the dataset.

**Ans:** Skewness of above boxplot is positive 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.

**Ans:** Both of these graphs are positive skewed and also outlier are also same which is 25.we can easily find the mode value in the histogram further median value can visualize from boxplot.

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** one in 200 long-distance telephone calls are getting misdirected.

Probability of call misdirecting   = p = 1/200

Probability of call not Misdirecting = q = 1-1/200 = 199/200

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

Number of Calls = n = 5

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

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

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

P(1) = 0.0245037

Therefore, Probability of 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 $2000 as for $2000, the probability is 0.3 which is high as compared to others.

1. Is the venture likely to be successful? Explain

**Ans:** Yes, 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

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

**Ans:** The long-term average earning of business ventures is expected value E(X)

E(x) =

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

= 800

Therefor, Average return 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 a venture depends on the Variability in the distribution. Higher Variance means more chances of risk

Var (X) = E(X2 ) –(E(X))2

= 2800000 – 8002

= 2160000

E(x2) =

= (-20002 ) \* 0.1 + (-10002) \*0.1 + 02\*0.2 + 10002\*0.2 + 20002\*0.3 + 30002\*0.1

= 2800000

**Variance of x is 2160000**