**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:**

**data={'Name of company':['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'],'MeasureX':[24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00]}**

**df=pd.DataFrame(data,columns=['Name of company','Measure X'])**

**df**

**Mean(Measure X) = 33.271333**

**Var (Measure X) = 287.146612**

**Std dev(Measure X)=16.945401**

Q1=np.percentile(df['Measure X'],25)

Q3=np.percentile(df['Measure X'],75)

IQR=Q3-Q1

lower\_wesker=Q1-(1.5\*IQR)

upper\_wesker=Q3+(1.5\*IQR)

upper\_outliers=df[df['Measure X']>upper\_wesker]

lower\_outliers=df[df['Measure X']<lower\_wesker]

outliers=[upper\_outliers,lower\_outliers]

outliers



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:(i)Q3-Q1=12-5=7 (It is equal to median of the data)**

**(ii)Given data is right-skewed**

**(iii)Only scale will change**



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(i)mode lies in interval [4,8]**

**(ii)Right-skewed**

**(iii)Our data lies between 0 to 20, and only one outlier(25).**

**Here mean>median, bimode,Right-skewed,less-picked**

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: p(misdirected call)=1/200**

**P(successive call)=1-(1/200)=0.995**

**Given each attempts are independent so**

**P(successive call after 5 attempt)=(0.995)\*( 0.995)\*( 0.995)\*( 0.995)\*( 0.995)=0.975**

**P(misdirected call after 5 attempt)=1-0.975=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?
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 : i) The most likely monetory outcome of the business venture is $2000 because it has highest probability occurrence .**

**ii) YES, we see that unique probability occurrence are 0.1,0.2,0.3**

**there fore 0.1+0.2+0.3=0.6\*100**

**=60% so the venture likely to be successful.**

**iii) consider similar business venture**

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

**=800**

**There fore the long-term average earning of business venture of this kind will be $800.**

**iv) X.std()= 1870.8286**

**X.var()=3500000.0**