**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)

>plt.boxplot(df[‘Measure\_X’],vert = False)



Outliers: Morgan Stanley 91.36%

df[‘Measure\_X’].describe()

0.3327133

df[‘Measure\_X’].var()

0.02871466

df[‘Measure\_X’].std()

0.169454



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) Inter-quartile range is the range between upper quartile (Q3) and lower quartile (Q1).

IQR= Q3-Q1= 12-5 = 7

50% of the data lies between IQR.

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

Ans) From the above boxplot we can say that the distribution of X is right-skewed or positively skewed.

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 it was found that the data point is actually 2.5 instead of 25, the outlier in the boxplot will be removed.

Whether the median shifts or not depends on the size of the data.

It will reduce the right skewness of the data.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans) The mode can lie between 4 and 10 because there are many values in this range but this is just an assumption.

1. Comment on the skewness of the dataset.

Ans) It is right skewed or +ve 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) from the above histogram and barplot we can confirm an outlier at 25 in Y value. Both the plots indicate the +ve skewness of the dataset.

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) Ans. Probability of wrong number P(W)= 1/200 = 0.005

Probability of not wrong number P(NW)= 1 – 0.005 = 0.995

Probability of at least 1 wrong number out of 5 calls = 1 – all 5 calls are not wrong number

= 1 – {1 - P(W)}^5

= 1 – 0.975

= 0.024

= 2% chance.

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) As the probability of X = 2000 is highest that is 0.3. hence it is most likely monetary outcome..

1. Is the venture likely to be successful? Explain

Ans) Yes, the venture is successful. For a successful venture the condition is that:

P(less than or equal to zero return) &lt; P(more than zero return)

And from given data the condition is satisfied.. (Note: 0.2+0.3+0.1=0.6=>0.6\*100=>60%).

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

Ans) (iii)long term average earning = expected value = (-2000)\*0.1+

(-1000)\*0.1+(0)\*0.2+(1000)\*0.2+(2000)\*0.3+(3000)\*0.1=800

Therefore the long-term average earning for these type of ventures would be around $800.

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

Ans) Standard deviation is the good measure of measure of the risk involved in this venture.

Std. deviation = sqrt((sq(summation of (X – x\_bar))/N)> df[“x”].std()

1870.829

> df[“x”].var()

3500000

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