**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% |

Answer:

Arranging the data in ascending order

24.23, 24.39, 25.41, 25.53, 25.81, 25.99, 26.71, 28.25, 29.62, 32.95, 35.00, 39.42, 40.26, 91.36

Find the median from the dataset

(26.71+28.25)/2

=27.48

Find the Quartiles

The first quartile is the median of the data points to the left of the median

24.23, 24.39, 25.41, 25.53, 25.81, 25.99, 26.71

Q1=25.53

The third quartile is the median of the data points to the right of the median

28.25,29.62,32.95,35.00,39.42,40.26,91.36

Q3=35.00

Find the minimum and maximum numbers

Minimum=24.23

Maximum=40.26

Find Outliers by using IQR formula

IQR=Q3-Q1

=9.47

Q1=Q1-1.5\*IQR

=25.53-1.5\*9.47

=11.325

Q3=Q2+1.5\*9.47

=35+1.5\*9.47]

=49.20

There is one outlier in the data 91.36 is above the maximum value

Mean= 33.92

Variance= 302.3627

Standard deviation=17.388



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?

Answer:

i. The IQR describes the middle 50% of values when ordered from lowest to highest. To find the interquartile range (IQR), ​first find the median (middle value) of the lower and upper half of the data. These values are quartile 1 (Q1) and quartile 3 (Q3). The IQR is the difference between Q3 and Q1.

Above boxplots IQR

IQR=13-5

=8

The IQR value 8 is not too for from the median

1. Given boxplot is right skewed hence we can say it’s a positive skewness
2. And we can say their will be no outliers in above boxplot



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.

Answers:

1. mode of this dataset lie on 20.
2. This distribution having less values on right side and more on left side , hence we can say it is an positive skewed data.
3. Both histogram and boxplot are used for visualization of data.it gives information about data. Box plot of a dataset gives minimum value, first quartile, median, third quartile, maximum value of the data. We can also find wheather there is an outlier or not specifically from boxblot.

In histogram we get idea about mode value, skewness etc. Histogram doesn’t provide idea about outliers. But it can show the distribution of data , how well a data got distributed .

So both plots have their own use and support on data visualization.

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.)

Answers:

Probability of misdirecting call,  p = 1/200

      Probability of not Misdirecting calls = 1 - 1/200 = 199/200

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

= 1 - Probability that no attempted call reaches the wrong number

=1-(199/200\*199/200\*199/200\*199/200\*199/200\*)

=0.025

**Probability that at least one in five attempted telephone calls reaches the wrong number**

**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

Answers:

1. The most likely monetary outcome is 2000 ,because it has highest probability
2. we can say that the venture is success. Basically, the total probability is 1. Among this, probability of positive values of x are 0.2+0.3+0.1=0.6 which is greater than the probability of negative values, which is 0.1+0.1=0.2 is likely to be not successful.
3. = (-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000\*0.3)+(3000\*0.1)

= -200-100+0+200+600+300 =1100-300

=800

iv. Variance is the good measure of the risk involved in the venture.

That is , Var(x)=216000.