**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:

Mathematically:

Mean=33.27,

Standard deviation=16.94,

Variance=287.14.

Now to calculate Outliers using formula.

Median(Q2)=26.71.

So, we will find out Q1 and Q3 to get our outliers using IQR.

sortedData=(24.14,24.23,24.39,25.41,25.53,25.81,25.99,26.71,28.25,29.62,32.95,35.00,39.42,40.26,91.36).

Qsort1=c(24.14,24.23,24.39,25.41,25.53,25.81,25.99);

Qsort2=c(28.25,29.62,32.95,35.00,39.42,40.26,91.36);

So, Q1= 25.41

Q3=35.

So, IQR=Q3-Q1=10.59

So outliers will be below Q1-1.5\*IQR value and above Q3+1.5\*IQR value.

Q1-1.5\*IQR= 25.41-(1.5\*10.59)= 9.525.

Q3+1.5\*IQR=25.41+(1.5\*10.59)=41.295.

So any value below 9.525 and above 41.295 is a outlier.

So, outlier=91.36.

R code

Ques1<-c(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)

Qsort<- sort(Ques1, decreasing = FALSE)

> hist(Qsort)

> boxplot(Qsort,horizontal=TRUE)

> outliers<- boxplot.stats(Qsort)$out

> View(outliers)

91.36



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: As per the boxplot,

Q3=12. Approx

Q1=5.

So, IQR-Q3-Q1=12-5=7.

IQR can be used as a measure of how spread-out the values are.

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

ANS: Data is positively skewed.(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?.

ANS: If we remove 25 and add 2.5, there will be no outliers in the dataset.

Also, Q1(Quartile 1) value will decrease by adding 2.5.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

ANS: Mode(Most frequent value) will lie in 4 to 8 Y values.

1. Comment on the skewness of the dataset.

ANS: Positively 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: As both Boxplot in Ques2 and Histogram in Ques3 is almost same.

-Both depicts that the given dataset has Outlier(25).

-Dataset is Positively Skewed.

-Most of the data is centered between (4,12).

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:

Probability of the event call misdirected is

P(E) = 1/200

So,

P(bar E)=1-P(E)=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)^5.

=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?

ANS: 2000.

1. Is the venture likely to be successful? Explain.

ANS: Venture will be successful as positive returns has probability 0.6 (0.2+0.3+0.1=0.6) i.e more than 50%.

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

ANS: Expected Value: 800.

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

ANS: Standard deviation.