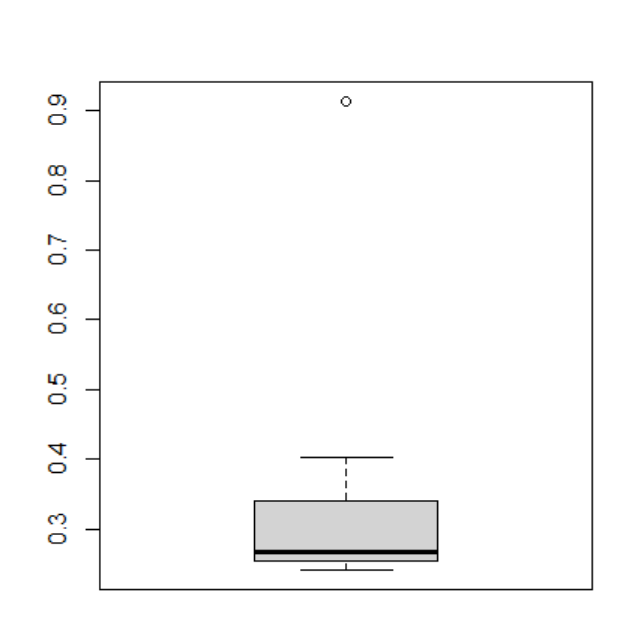
**Topics: Descriptive Statistics and Probability**

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:



Outlier value is: 0.9136

0.3327133

= 0.169454

= 0.02871466



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 = 12.5-5 = 7.5

It explains the spread of the data in the middle half ie, 50 % of data lies within Q1 to Q3 .

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

Ans: This data is positively skewed. The median comes closer to the whisker. Mean> median.

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: As the outlier values are -6.25 and 23.75 beyond lower and upper whiskers respectively. If we remove 25 and place 2.5, the outlier value in this data set will be removed. Also, mean and median values get lowered accordingly. Then Q1 and Q3 will get changed. Skewness will not get affected as it is already positively skewed.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: Mode of the data set located between 4 and 8.

1. Comment on the skewness of the dataset.

Ans: The data is right/positively skewed, only few data are leading to the right end of the histogram.

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: Boxplot is giving info. related to the position of median and most data (50%) occurring range. It is suitable and more explanatory if data count is less. Histogram shows the frequency of data and give info. about mode value location. Both data shows the outlier information. Skewness info can be interpreted with these. Histogram is explaining the frequency and distribution of data and more suitable for large data sets.

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: number of calls, n= 5

Given data: one in 200 long-distance telephone calls is misdirected.

Hence, Probability of call misdirection, p= 1/200

Probability of call not misdirecting, p = 1-1/200 = 199/200

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

at least one call out of 5 reaches wrong number means chances are for:

1call ,2 calls,3 calls,4 calls,5 calls gone wrong : P(1),P(2),P(3), P(4), P(5)

zero calls chance is excluded.

Hence probability for at least 1 call to get misdirected will be: 1- P(0)

= 1- 5C0p0q5-0

= 1- (199/200)5

= 0.02475

Probability for at least one call get misdirected is = 0.02475

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: x= 2000, because it has the highest probability of 0.3

1. Is the venture likely to be successful? Explain

Ans: The venture is likely to be successful, because P(x=1000)+P(x=2000)+P(x=3000) = 0.2+0.3+0.1 = 0.6

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

Ans: Long-term average earning is the expected value, E(X) = ∑x.P(x)

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

= 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 to evaluate the risk.

Sd = 1870.829