**Topics: Descriptive Statistics and Probability**

* Look at the data given below. Plot the data, find the outliers and find out

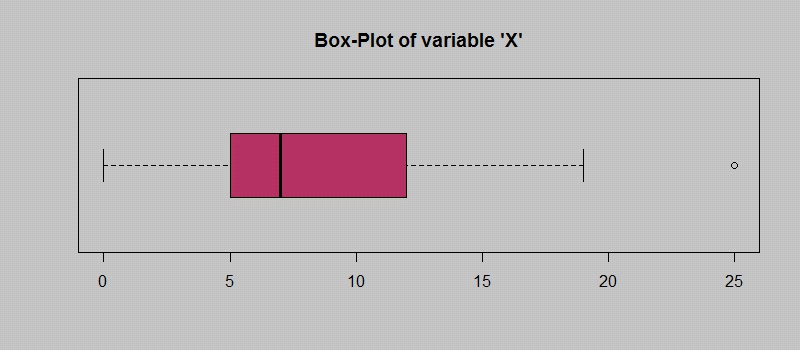
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

Ans: **the outlier according to data is:**

|  |  |
| --- | --- |
| **Morgan Stanley** | **91.36%** |

**all the other values are below 41%.**

**IQR=0.08505 and as morgan Stanleys 0.9136 > IQR, it is an outlier**



Answer the following three questions based on the box-plot above.

* 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)=7 This value implies maximum data points are concentrated in this range.**

**The first quartile (Q1) is median of first n values=Q1=5**

**The third quartile (Q3) is the median of n largest values = 12**

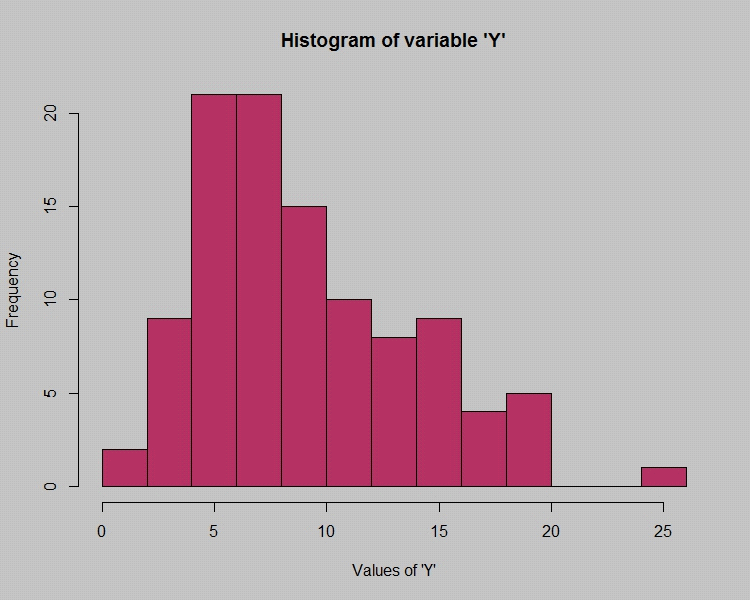
* What can we say about the skewness of this dataset?

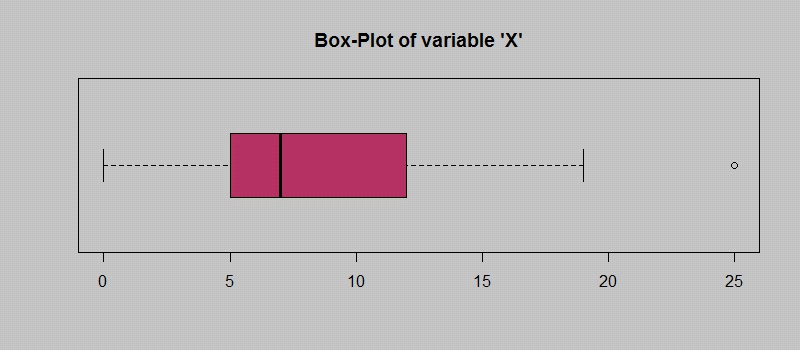
Skewness= **The length of the right whisker is greater than the left whisker. From this, we can conclude that the data is positively skewed. Therefore, mode < median < mean.**

**Q3-Q2 > Q2-Q1**

* If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

**The Q1 and Q2 i.r the first and second quartile will get changed slightly and shift towards left but the Q3 will probabbly stay the same.**





Answer the following three questions based on the histogram above.

* Where would the mode of this dataset lie?

Ans: Right Skewed data: **mode<median<mean, mode will lie of the left side(and it will be the least amongst all 3)**

* Comment on the skewness of the dataset.

Right Skewed data: **mode<median<mean. most of the data points are on left and outliers are present on the right side of the distribution. Thus it is right skewed dataset or positively skewed dataset.**

* 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.

Conclusion from observing the above graphs:

Median: **Q2: it lies between 5 to 10 i.e. around 7 and it is on the left side of the mean.**

**outlier is at 25**

**Range of the data spread is between 0 to 20**

**50% of the data lies in IQR i.e. from 5 to 12(between these 7 datapoints thus, IQR is 7)**

**As the data is rightly skewed, as the tail of both boxplot and histogram extends towards right side, it is positively skewed, representing many data points lie in the right side if the median.**

**mean will also lie on the right side region.**

* 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: 0.025**

* 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 |

* What is the most likely monetary outcome of the business venture?

**Ans: most likely monetary outcome of the business venture  is 2000 $ as it has maximum probability = 0.3**

* Is the venture likely to be successful? Explain

**Ans: it will be almost successful as the expected value is 800$ .**

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

**800.0 (Expected value)= E (xi\*pi)**

* What is the good measure of the risk involved in a venture of this kind? Compute this measure: