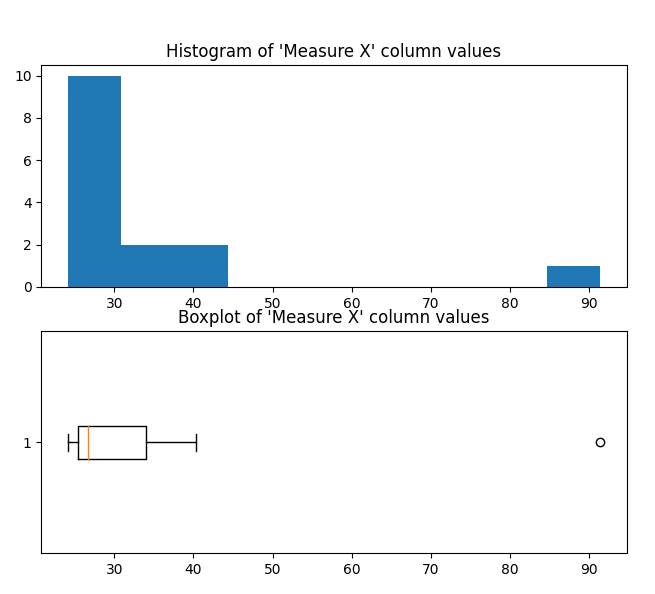
**Topics: Descriptive Statistics and Probability**

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% |

**Histogram, Boxplot of ‘Measure X’:**



**From the above histogram, boxplot, we can say that the data point “Morgan Stanley” is an outlier.**

**Mean(**μ **):** **33.27133**

**Variance(**σ2 **):** **287.1466**

**Standard Deviation(**σ**):** **16.9454**

**Respective Code files: Set1\_Q1.R, Set1\_Q1.py**



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 region is where 50% data points lie. It is the difference between 75%, 25% quartiles. Here, Q1=5, Q3=12 approximately. So, IQR = Q3-Q1 = 7.

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

Ans: Most of the data is lying to the left of the plot(has a tail to its right). So this data set is right-skewed.

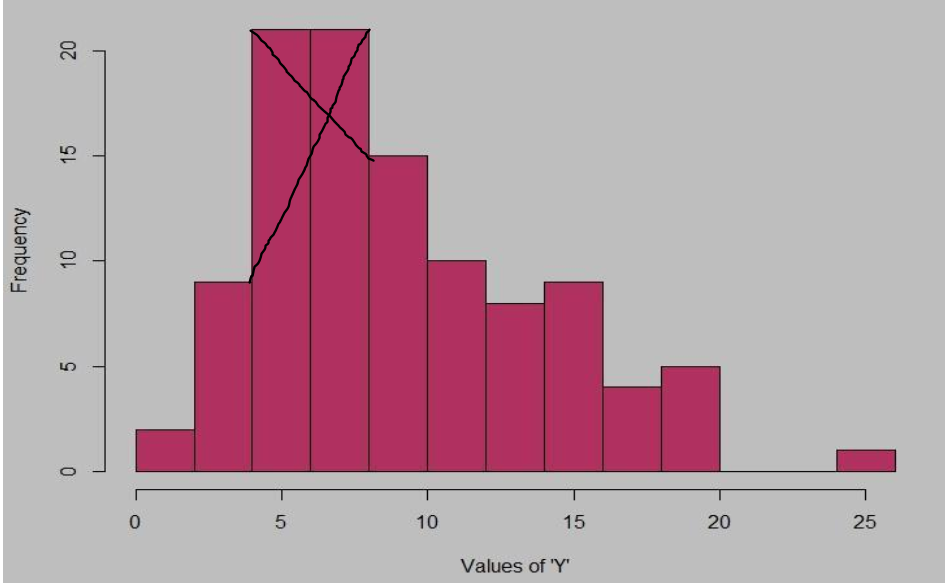
1. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

Ans: The outlier would be removed and, 2.5 being the lowest value, IQR would be 12-2.5 = 9.5



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: 

Mode can lie anywhere between 4 to 10

1. Comment on the skewness of the dataset.

Ans: Right 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: Based on the data distribution, both the plots show the data is Right Skewed. Data has an outlier 25.

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: From the given data, one out of 200 calls is misdirected. So, the probability of a call misdirecting = 1/200

probability of a call not misdirecting is 1-1/200 = 199/200

Probability that at least one in 5 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 as it has the highest probability 0.3

1. Is the venture likely to be successful? Explain

Ans: Yes. 60% of the time they got profit.

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

Expected value (EV) is an anticipated value for an investment at some point in the future

EV = 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: (sd(x)) = 1870.829