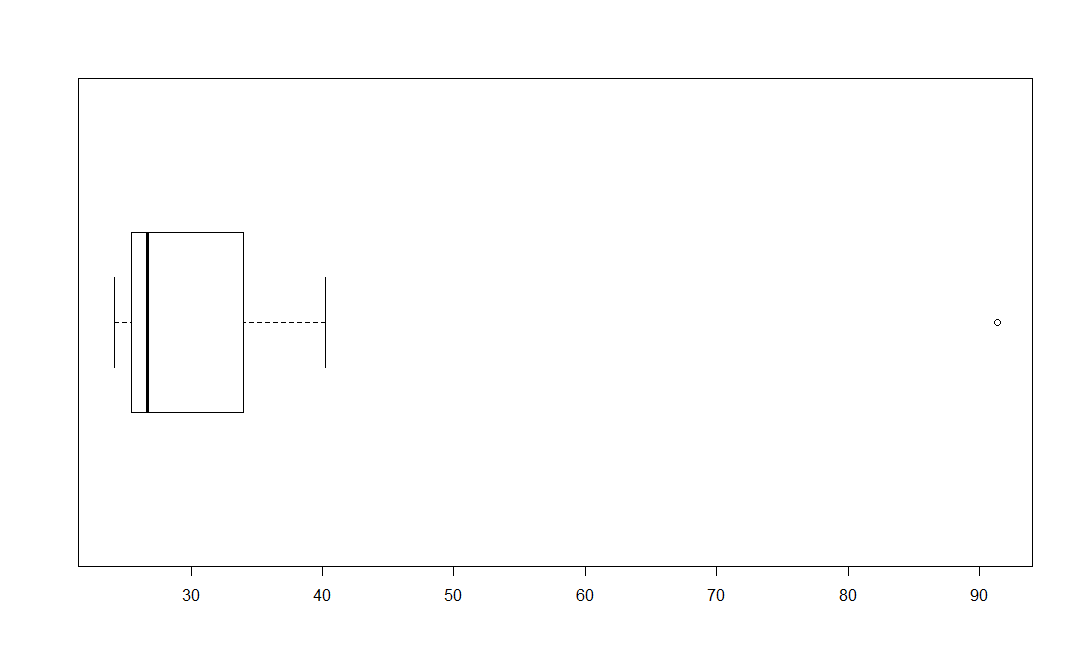
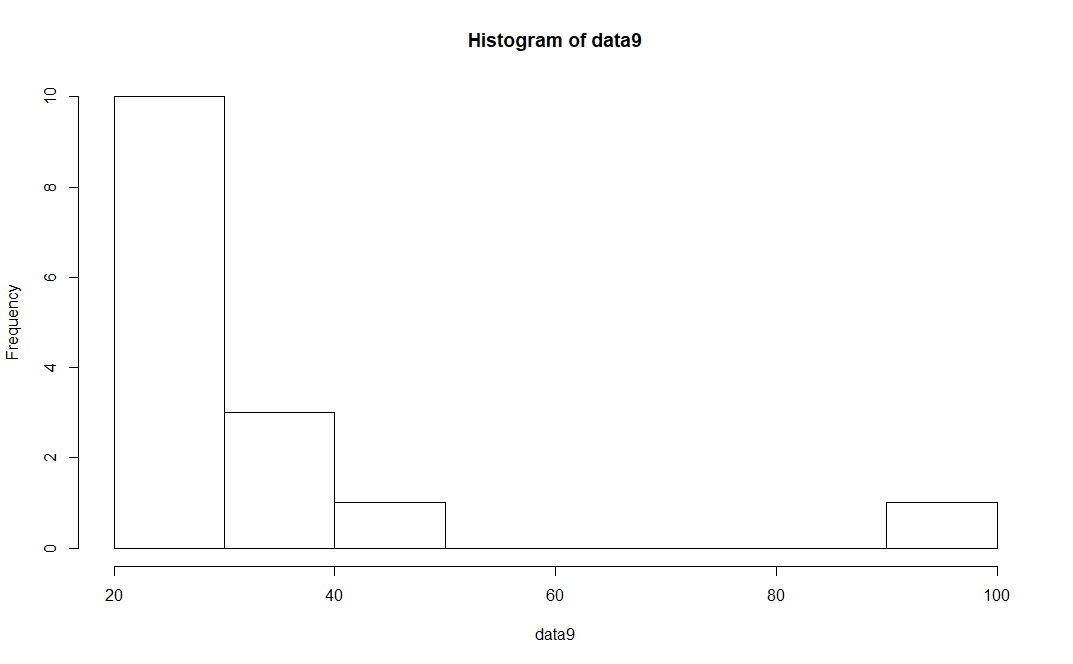
**Topics: Descriptive Statistics and Probability**

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% |

* Mean = 33.27
* Variance = 287.14
* SD = 16.94





By looking at the above 2 visualizations, we can say that the data point “91.36%” is the outlier.



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. Q3 – Q1 = 12 – 7 = 5 (50% of the data points lie in IQR)
2. What can we say about the skewness of this dataset? Right Skewed
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? Skewness would be reduced. There will be no outliers in the data



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie? Mode may lie between 4 to 8 looking at the data. Need to have data to calculate actual mode value.
2. Comment on the skewness of the dataset. Right Skewed
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.

* Both diagrams are plotted using same data set
* Box-plot tells us about lower & upper extremes, Q1, Q3, Median and skewness of the data
* We can’t comment about mode using a box-plot
* Histogram also tells us about the skewness of the data
* Using histogram, we can say where the mode of the data lies, but need to have original data to find out mode

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.)

Probability of wrong phone calls = 1/200

Probability of right phone calls = 1-1/200 = 199/200

Probability of at least 1 in 5 calls reach wrong number = 1 – Probability of none of the 5 calls reach wrong number

* 1 – Probability of 5 right phone calls
* 1 – (199/200)^5 [Since 5 event are independent of each other]
* 1 – (0.995)^5
* 1 – (0.975249)
* 0.024751
* 2.48%

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? 2000 (30% is the highest probability of occurrence)
2. Is the venture likely to be successful? Explain

Probability of Positive Outcomes = 0.2+0.3+0.1 = 0.6

* 60% chance to be successful

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

Expected Value = ∑ 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

Variation is a risk in any kind of business. The less the variation, the better would be the outcome of any business

Variance = σ2 = E[(X - µ)2] = ∑ (X - µ)2 P(X) = (-2000-800)2\*0.1 + (-1000-800)2\*0.1 + (0-800)2\*0.2 + (1000-800)2\*0.2 + (2000-800)2\*0.3 + (3000-800)2\*0.1 = 2160000

* SD = SQRT (Variance) = 1469.693846
* SD > Mean
* This business is risky