**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% |

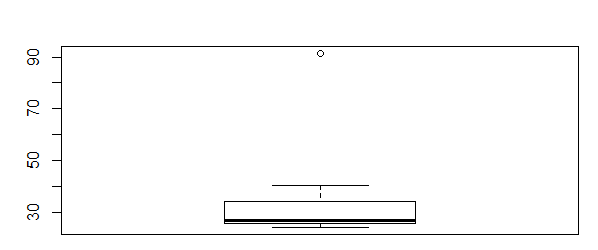
Answer:

***µ = 33.27133***

***σ = 16.9454***

***σ2 = 287.1466***

***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.

Answer: ***IQR = Q3-Q1 = 12 – 5 = 7 (approximately). It means 50% of data points lie in the range of 5 and 12***

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

Answer: ***It is positive/left skewed as longer tail on right side***

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?

Answer: ***Inner quartile range will change and no outlier will be present***



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans:- ***The mode lie in between 4 and 7***

1. Comment on the skewness of the dataset.

Ans: ***Its positive/left skewed as it has longer tail on right side.***

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***: • The histogram and box plot will help to understand the distribution of data sets.***

***• From the histogram it can be understand, how much the datasets is skewed and how higher the peak of the curve is.***

***• From the box plot, it can be helpful to identify the outliers in the datasets. Even the alignment of box in boxplot will help to understand the range in which maximum datasets lie and help us to identify the median through which we can understand the distribution of data above median and below median.***

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.)

Answer:

***One wrong call out of 200***

***Probability of wrong call = 1/200 = 0.005***

***Probability of not wrong call = 1-0.005 = 0.995***

***Probability of at least one out of five is a wrong number***

***= 1- Probability of all five calls are not wrong numbers***

***= 1- 0.995^5***

***= 0.0247***

***=2.5%***

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?

Answer: ***2000***

1. Is the venture likely to be successful? Explain

Answer: ***yes , there are 60% chances of getting a positive return and only 20% of a negative return.***

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

Answer***: for long term= -2000\*1 + (-1000\*1)+1000\*2+2000\*3+3000\*1 /6 = 8000/6=1333***

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

Answer: ***The good measure of risk is the standard deviation***

***sd(x) = $1870.829***