**Topics: Descriptive Statistics and Probability**

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

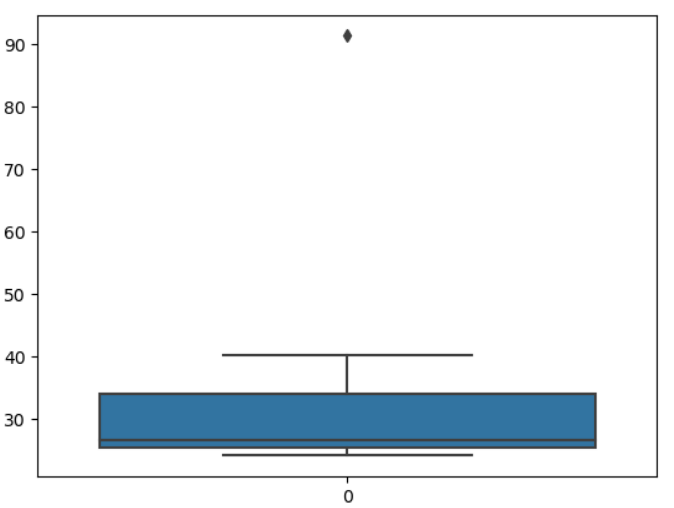
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft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

Ans: For the given data the outlier is 90.

Mean = 33.271

Standard deviation =16.945

Variance = 287.14





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: Approximately (first Quartile Range) Q1 = 5,Q3 = 12,Median (Second Quartile Range) = 7

IQR = Q3 - Q1

= 12 - 5

= 7

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

Ans: The skewness of the data 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: In that case there would be no outlier in the above box plot because of the outlier the data had positive skewness it will reduce and the data will have normal distribution.

-

Answer the following three questions based on the histogram above.

1. Where would the mode of this data set lie?

Ans: The mode of this data set would lie approximately between 4 to 8.

1. Comment on the skewness of the data set.

Ans: The skewness of the data is positively skewed.

1. Suppose that the above histogram and the box-plot in question 2 are plotted for the same data set. Explain how these graphs complement each other in providing information about any data set.

Ans: They both are positively skewed and both have outliers.The median can be easily visualized in the box plot where as in histogram mode is more visible.

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: The probability that at least one in five attempted telephone calls reaches the wrong numbers is approximately 0.9950% which is 1%.

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: The most likely monetary outcome of the business venture is 2000$ as the probability is 0.3 which is maximum as compared to others.

1. Is the venture likely to be successful? Explain

Ans: Venture is successful if X is +ve Hence if Xis 1000,2000 or 3000.Probability is 0.2+0.3+0.1=0.6 as 0.6>0.5 Hence venture likely to be successful

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

Ans: The long-term average is the Expected value = SUM(X \* P(X))

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

= -200 -100 +0+200+600+300

= 800

Therefore the long-term average earnings of business ventures of this kind is $800.

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

Ans: The good measure of the risk involved in a venture of this kind depends on the variability in the distribution.

Higher Variance means more chances of risk,

Var(X) = E(X^2) - (E(X))^2

= 2800000 - 800^2

= 2160000

SD=Var ≈1470

As variability is quite high Hence risk is high