**Topics: Descriptive Statistics and Probability**

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

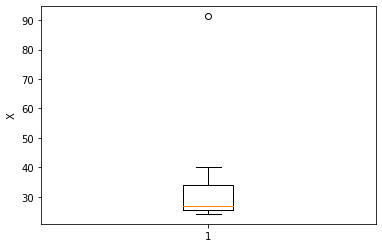
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan& Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft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

**Ans:**

-Mean (µ)= 33.27133

-Std. Deviation (σ)= 16.3708

-Variance (σ^2)=268.00350

-Outlier= Morgan Stanley(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.**

**Ans:** IQR of the data set is approximately between 5 to 12. This range implies that the major

part of data lies between the this range along with the median approximately at 7.

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

**Ans:** The plot seems to be left skewed, which is positively 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:** Firstly the data point which is considered as the outlier will be now included in data

points lying on the whisker. Also the upper and lower extremes of the box plot whisker

may change

****

**Answer the following three questions based on the histogram above.**

1. **Where would the mode of this dataset lie?**

**Ans:** The mode of the data set would lie between 5 to 8.

1. **Comment on the skewness of the dataset.**

**Ans:** The Histogram seems to be left skewed, which is positively 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:** While the histogram representing the frequency distribution of the data, on the other hand the boxplot is giving the detail about how the data is distributed in certain range. It displays two common measures of variability in the data set

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:**

one in 200 long-distance telephone calls is misdirected

=> probability of call misdirecting p = 1/200

Probability of call not Misdirecting = 1 - 1/200 = 199/200

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

at least one in five attempted telephone calls reaches the wrong number

= 1 - none of the call reaches the wrong number

= 1 - P(0)

= 1 - ⁵C₀(1/200) ^0 (199/200)⁵⁻⁰

= 1 - (199/200)⁵

= 0.02475