**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% |

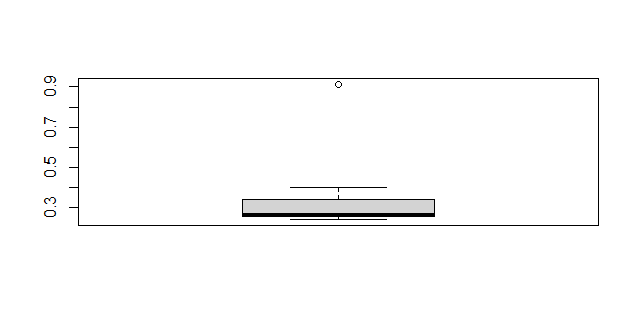
As per above data set and as per attached R file with firm.xl dataset excel file Please found the following values which are following:

Mean (**μ) =** 0.3327133

Standard Deviation **(σ) =** 0.169454

Variance **(σ^2) =** 0.02871466

Also see the below Box-plot and as per attached R file with firm.xl data set only one outlier is available.

  
Outlier details are following:

Morgan Stanley = **0.9136 or 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:** Inter-quartile range of this dataset is = 7 & this value shows the 75% value of dataset or

range between upper & lower quartile.

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

**Answer:** Skewness of this dataset is negative due to median point is more close to the lower

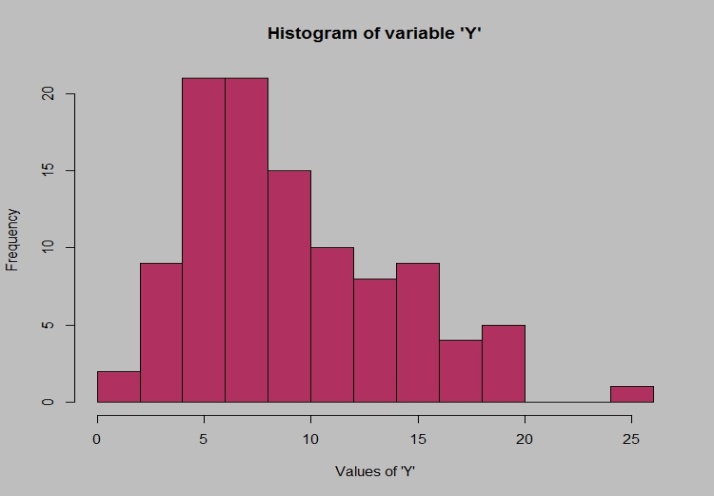
quartile.

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:** First value 25 is outlier in above dataset box-plot & if we plot new box for this dataset

with 2.5 value instead of 25 then box plot will cover all data & there would be no

outlier in new box-plot.





Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Answer:**  Mod of this dataset lie between 4 & 10 data points on Y axis.

1. Comment on the skewness of the dataset.

**Answer:** Skewness of the above dataset is negative.

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.

**Answer:** As per above dataset & question 2 box plot both have so many difference about providing

information which are following:

a] Histogram represents frequency distribution for numeric data & Box plot summarizes the important

aspects of the distribution of continuous data.

b] Both plot are providing the data in box with lower, upper quartile, median, minimum & maximum

value. Histogram represents only maximum & minimum value with nature of data.

c] Histogram allows to visually and quickly accesses the shape of the distribution, the central tendency,

the amount of variation in the data, and the presence of gaps, outliers or unusual data points.

Box-Plot provides a quick way for examining the variation present in the data.

d] Both-plot provides the skewness, data nature such as symmetric or asymmetric shape of

data structure.

e] Box-plot needs small space & Histogram needs more space to provide information about dataset.

f] Box-plot contains five values need to calculate from dataset to construct itself in box & Histogram

needs only all data points to construct itself from dataset.

g] Box-plot shows clearly outliers in its structure But for histogram we need to study & then confirm the

outliers from dataset graphical representation.

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: Probability of one in 200 long-distance telephone misdirected call= 1/200 = 0.005

Probability of one in 200 long-distance telephone not misdirected call = 1-1/200

= 199/200

= 0.995

No. of calls = 5

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

= 1 – none of the calls reaches wrong number

= 1 – (1- 0.005) ^5

= 0.024

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: Monetary outcome of the business venture is 2000 because it has highest probability

0.3.

1. Is the venture likely to be successful? Explain

Answer: Venture is successful because in business venture probability (0.2, 0.3 &0.1) for

1000, 2000, 3000 is very high as compare to expected loss probabilities (0.1 & 0.1)

for only -2000 & -1000.

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

Answer: The long term average earning of business venture of this kind is following:

n

Long Term Average Earning = ∑ xp(x)

i=1

= (-2000)\*0.1 + (-1000)\*0.1 + 0\*0.2 + (1000)\*0.2 + 0\*0.1 +

(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

Answer: The good measure of the risk involved in this kind of business venture is standard

deviation because by standard deviation we get to know that how-much data is varying

from its mean for risk involvement in business and its also used to make predictions.

Mean(μ) = -2000-1000+0+1000+2000+3000/6

= 500

Standard Deviation (**σ)** = ∑ sqrt(Xi – μ)^2/N

= sqrt[(-2000 - 500)^2 + (-1000 - 500)^2 + (0 - 500)^2 +

(1000 - 500)^2 + (2000 - 500)^2 + (3000 - 500)^2]/6

= sqrt[ (-2500)^2 + (-1500)^2 + (-500)^2 + 500^2 + 1500^2+ 2500^2]/6

= sqrt[ 6250000 + 2250000 + 250000 + 250000 + 2250000 + 6250000]/6

= sqrt[2916666.6667]

= 1707.825

Also find the computation of standard deviation in attached R File with this question business venture data excel file.