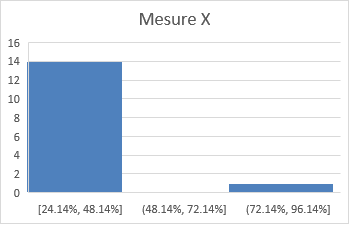
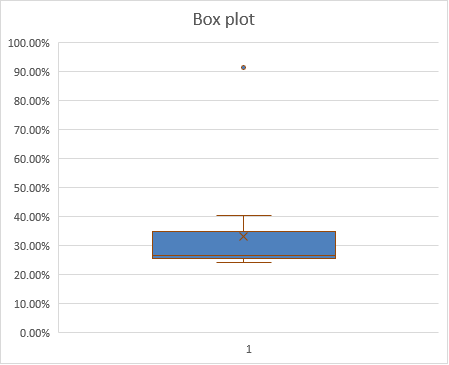
**Topics: Descriptive Statistics and Probability**

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 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. *Inter quartile range is (12-5=7). It tells about the spread of middle half of the data set. Large Inter quartile range depict that the central portion of the data spread out further. And small inter Quartile range tells that middle values cluster more tightly.*

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

*Ans. From the box plot we can infer that the median is towards the left which implies that the data is right Skewed. That is data points are concentrated towards the left-hand 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?

Ans. If the value 25 (which is currently an outlier for the current data set) is replaced with 2.5 which help to widen the IQR slightly towards the left ,and the outlier will be removed. And 2.5 falls in the trend of data set.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans. *The mode which mean the most repeated no of data point. here in this question 2 values are having same frequency so both the values are considered to be mode values.*

1. Comment on the skewness of the dataset.

Ans. *For the above histogram we can infer that the histogram is lightly tailed towards right side. Which means the histogram is right skewed or 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. The Box plot

* gives us more information like outliers and their values ,the middle value between median and the smallest number(q1-first quartile).middle value between median and the highest value of the data set.
* And the box plot also helps gives us information like inter quartile range (gives us the idea where the middle 50% of the data lies.
* Gives us the idea about the consistence of the data points.)
* Helps to where the median of the data set.
* Helps easily find which is minimum and maximum values of the data points.

In case of Histogram:

* It helps to easily interpret the frequency of different values range.
* In box plot gives the median of the data points here in case of histogram gives the mode of the data set.
* It also gives the idea about the outliers and their frequency.

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.)

This is can considered as case of binomial distribution

P(misdirected)=1/200

P(not misdirected)=199/200

n=5

p=1/200

q=199/200

p(atleast 1 call get misdirected)=1-p(none of the call reach wrong number)

1. [

=1-

=1- 0.975248753121875

= 0.02475

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?

The most likely monetary outcome of the business venture is 2000 as it is having the max probability (.3).

1. Is the venture likely to be successful? Explain.

For the venture to be success x>0

Probability of (x>0) = p(1000)+p(2000)+p(3000)

=.2 + .3 + .1

= .6

Yes it has 60% probability of being successful.

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

The long term average earning can be calculated by calculating the expectation

E(X)=∑x\*P(X) =800

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

The measure risk involved can be calculated based on the standard deviation

E()=2,800,000

Variance=21,60,000

Standard deviation=1,469.69

