**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:

Measure = pd.read\_excel('/Users/Akshay/Documents/DataScience/ExcelR/Assignments/DataScience/Basic Statistics\_Level-2/Assignmnt\_q.xlsx')

measure.MeasureX.describe()

count 15.000000

mean 0.332713

std 0.169454

min 0.241400

25% 0.254700

50% 0.267100

75% 0.339750

max 0.913600

Name: MeasureX, dtype: float64

measure.MeasureX.var()

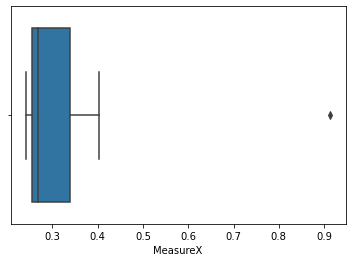
0.028714661238095233

Mean : 33.27%

Std Deviation : 16.94%

Variance : 2.87%

To find outliers: box plot



0.91 is the point which is away from the upper extreme points i.e Morgan Stanley



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.

IQR= Upper Quartile – Lower Quartile

=12-5

=7--------------IQR

50% of data points lies in the IQR

IQR lies from 5 to 12

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

The dataset is positive skewed as the right whisker length is more and also the upper IQR is far away from the median

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?

The new box plot will not be having any outliers if the value 25 is actually 2.5



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Mode: Values between 4-8

1. Comment on the skewness of the dataset.

The dataset is positive skewed distribution as there is more tail towards +ve/right side of the graph

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.

Box plot would give us quick results of outlier where as

in histogram points the dense area like where most of the values fall(frequency) along with skewness of data and also it will be ease to find the mode just by looking histogram graph

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:

Probability of call misdirected , p = 1/200

Probability of call not misdirected , q = 1-(1/200)

n=5

Assuming no call reaches wrong number

Means r =0

=nCrp^rq^(n-r)

=5C0(1/200)^0(199/200)^5

5C0 = 5!/(0!\*(5-0)!)= 5!/5!=1

(1/200)^0 = 1

=1\*1\*(199/200)^5

=0.97525

At least one in five attempted telephone calls is misdirected: means 1 or more

=1-no call reaches wrong number

=1-0.97525

=.02475

2.475% probability is there that the call might misdirect among 5 attempted call in 200 distance

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 means highest profit maid

$2000 as the P($2000) is 0.3 which is higher than the other outcomes.

1. Is the venture likely to be successful? Explain

Expected value=x\*P(x)

=-2000\*.1+-1000\*.1+0\*.2+1000\*.2+2000\*.3+3000\*.1

= 800

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

long-term average earning of business ventures = $ 800

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

There is no risk involved as the expected value is +ve i.e $800