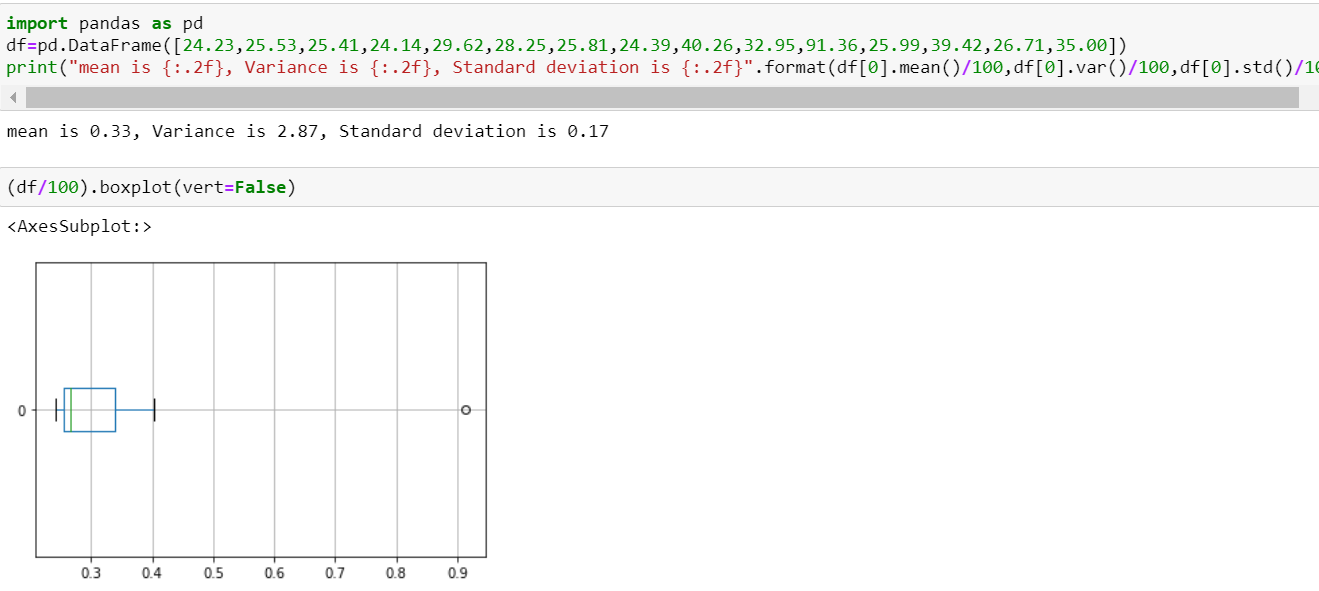
**Topics: Descriptive Statistics and Probability**

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% |

**Solution:**





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.
2. What can we say about the skewness of this dataset?
3. 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:** 1) Q1=5, Q3=12 then IQR=Q3-Q1=12-5=7

2) Data is positive or right-hand side skewed.

3) If it was found that the data point is actually 2.5 instead of 25, the outlier in the boxplot will be removed.

Whether the median shifts or not depends on the size of the data.

It will reduce the right skewness of the data.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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:** - 1) 4-8 in this range the dataset lies in mode

2)The data are positive or right-hand side skewed

3) Histogram can well define the mode, frequency of dataset & boxplot well defines the median, range & outlier of the data but skewness is well understood by the histogram.

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) = nCx \* p^x \* Q^n-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 =2% chance

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?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

Ans: -1) The most likely outcome of this business venture is a return of $2000 as it has the highest probability of occurrence.

2) Yes, we will calculate the average the whole dataset =-2000\*0.1- 1000\*0.1+0.2\*0+1000\*0.1+2000\*0.3+3000\*0.1=$700 & we can say the 60% chance of venture likely to be successful.

3) long-term average earning for these types of ventures would be =-2000\*0.1-1000\*0.1+0.2\*0+1000\*0.2+2000\*0.3+3000\*0.1=$800

4) Risk stems from the possible variability in the expected returns. Therefore, a good measure to evaluate the risk for a venture of this kind would be variance or standard deviation of the variable X.

> sd(ex$x)

[1] 1870.829

> var(ex$x)

[1] 3500000