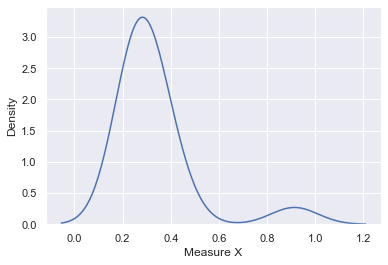
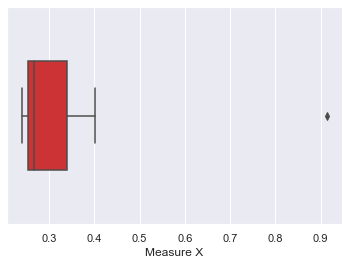
**Topics: Descriptive Statistics and Probability**

Note by Kripa – The Python notebook containing codes for the exercise is embedded here. 

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 1**



Mean = 0.332713

SD = 0.169454

Variance = 0.028715

Outliers

Name of company Measure X10

Morgan Stanley 0.913



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.

Q1 = 5

Q3 = 12

IQR = Q3-Q1 = 12 - 5 = 7

This implies that 50% of the data is concentrated between 5 and 12.

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

The data is skewed positively as median is closer to the left and whisker on the left is shorter.

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?

There will be no material impact on the box plot. The outlier 25 will be subsumed in Q1 if it becomes 2.5. Since minimum if 0 and Q1 is 5 there will not be significant change in left whisker.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

The mode will lie between 5 and 6.67 (of values of Y)

1. Comment on the skewness of the dataset.

The data is positively skewed i.e right tailed.

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.

It can be quickly interpreted from the histogram that the data set is not normally disributed and is skewed towards right. The variability demonstrated by boxplot (IQR) can be easily visualised with the help of 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.)

Given that 1 in 200 calls is misdirected. So probability of wrong numbers is

p(wrong) = 1/200 = 0.005

p(not wrong) = 1 – 0.005 = 0.995

To calculate the probability of atleast 1 in 5 calls reaches wrong number.

So probability of 1 or 2 or 3 or 4 or 5 wrongs needs to be calcualted. Which is equivalent to 1 minus probability of none of the calls reaching the wrong number. (AND condition)

So the probability of atleast 1 in 5 calls reaches wrong number = 1 – probability of all 5 reaching the right (not wrong ) number

So the probability of atleast 1 in 5 calls reaches wrong number = 1 – (0.995)^5 = 0.025 = 2.5%

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?

$2000 is most likely as the probability for 2000 is highest (0.3) in the distribution.

1. Is the venture likely to be successful? Explain

Yes as the probability of returns greater than 0 is 0.6.

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

Long term average earning is the expected value of the business. Expected value is calculated below –

|  |  |  |
| --- | --- | --- |
| **x** | **P(x)** | **Expected Value E(x)** |
| **x\* P(x)** |
| -2,000 | 0.1 | -200 |
| -1,000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 3000 | 0.1 | 300 |
| **Expected Value (Sum of individual expected values)** | | **800** |

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

A good measure of risk is standard deviation (SD).

(SD)2 = V(X) = E(X2) - [E(X)]2

|  |  |  |
| --- | --- | --- |
| **x2** | **P(x)** | **Expected Value E(x)** |
| **x\* P(x)** |
| 4,000,000 | 0.1 | 400000 |
| 1,000,000 | 0.1 | 100000 |
| 0 | 0.2 | 0 |
| 1,000,000 | 0.2 | 200000 |
| 4,000,000 | 0.3 | 1200000 |
| 9,000,000 | 0.1 | 900000 |
| **Expected Value of X2** | | **2800000** |

Hence SD = Sqrt (2800000 – 8002) = 1469.69