**Topics: Descriptive Statistics and Probability**

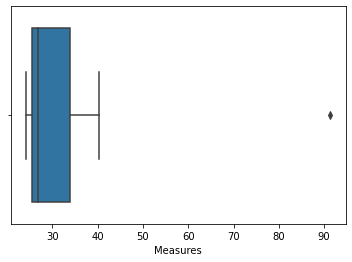
1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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% |

New\_data.mean() : 33.27%

New\_data.var() : 2.8714%

New\_data.std() : 16.94%



OUTLIER = print(np.where(New\_data['Measures']>50)) = Morgan Stanley(array 10)



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.

IQL range for this box plot is from {5,12}

Median value is 7

Min score is 0

Max score excluding outlier is 18

25% data lies below 5

25% data lies above 12

There exists one outlier in this box plot: 25

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

The plot is positively skewed, whisker is shorter on lower end

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?

Box plot will usually neglect the outliers, so even if outlier value is replaced there wont be significant effect on box plot



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Mode will lie between 4 to 8

1. Comment on the skewness of the dataset.

The ND is positively skewed. The data is more towards the lower side. That means there are more or less homogenous types of groups.

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.

Considering both the graph. The data can be declared as positively skewed. Almost 50% of data lies within IQR. 25 is outlier in this data.

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.)

Number of calls wrong = 1/200 = 0.005

Required call mistake ratio = 1/5

probability that at least one in five attempted telephone calls reaches the wrong number

= 1/200 \* 1/5 = 1/1000 = 0.001

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?

Most likely is highest P(x) = P(2000) = 0.3

1. Is the venture likely to be successful? Explain

P(1000)+P(2000)+P(3000) = 0.2+0.3+0.1 = 0.6

Therefore venture return is 60% successful, therefore venture is sucessfull

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

Earnings = (0.1)(−2,000) + (0.1)(−1,000) + 0 + (0.2)(1,000) + (0.3)(1,000) + (0,1)(3,000)

= 800

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

Risk measure = P(-1000) + P(-2000) = 0.1 + 0.1 = 0.2