**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: the mean is 33.27, standard deviation is 16.94 and variance is 287.14, there is only single outlier ie morgan stanley 91.36%.



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: the inter quartile range is 7, this values tells 50% of data points in the range of 5 and 12

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

Ans: since IQR is positive the data is positively skewed,tail is exending towards the right

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: the median would have remained the same and but we would have a different IQR range and we would see no outliers.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans:Mode is the smallest value and it is located at the left side of the dataset

Approximately between 4-8

1. Comment on the skewness of the dataset.

Ans: the skewness of this dataset is positively skewed and distributed towards the right

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: both the plots will be right skewed and would show us an extreme value as an outlier in the given dataset. the median can be seen in boxplot while in histogram we can visualise the mode.

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: the probability of reaching one in five attempted calls reaches the wrong number is 0.024 or .2%

#probability of calls getting misdirected

1/200

#probability of calls not getting misdirected

199/200

#since one in 5 attempts the call reaches the wrong number

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

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?

Ans: the most likely monetary outcome of this business is 2000$ as it has maximum probability is ie 0.3

1. Is the venture likely to be successful? Explain

Ans: yes this venture is likely to be successful, the probability of making greater is 79% as the probability is higher than 0.5 ie 0.79 .

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

Ans: the longterm average earning of buisness ventures is 800$ as we multiply the of x values to the p values and sum it up .

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

Ans: the measure of the risk depends on the variability of the distribution ,higher the variance more chances of risk , since the variance is higher the risk is higher

#computing the risk factor by variance

(-2000\*-200)+(-1000\*-100)+(0\*0)+(1000\*200)+(6000\*200)+(3000\*300)

#VAR(X)= E(X2)- ({E(X)})2

2800000-640000

#we know that standard deviation is square root of variance

np.sqrt(2160000)