**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% |
| JPMorgan & 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: -

* Mean = 33.27133
* Variance = 287.1466
* Standard deviation = 16.9454
* Morgan Stanley is the outlier in the Boxplot of 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: -

* IQR = 12-5 = 7, this represents the range which contains 50% of the data points and it have 1 outlier

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

Ans: -

* The skewness of this dataset is right skewed /positively skewed
* More than 50% of the data is between 7-12

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: -

* + will be not considered an outlier. The boxplot will start from 0 and send at 20 in representation.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: -

* Mode will lie between 4-8

1. Comment on the skewness of the dataset.

Ans: -

* The dataset is right skewed /positively skewed and also have outlier.

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: -

* Median in boxplot and Mode in histogram
* Histogram provides the frequency distribution so we can see how many times each data point is occurring however boxplot provides the quantile distribution i.e.,50% data lies between 5 and 12.
* Boxplot provides whisker length to identify outliers, no information from histogram. We can only guess looking at the gap that 25 may be an outlier.

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: -

* n=No. of calls =5
* P=probability of calls misdirects=1/200
* Q=probability of calls not misdirect=1-1/200=199/200
* at least one in five attempted telephone calls reaches the wrong number

nCr PrQn-r=1-none of calls reaches the wrong number

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

=1-(199/200)5

=0.02475

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: -

* Max. P(x) = 0.3 for x (2000). So most likely outcome is 2000

1. Is the venture likely to be successful? Explain

Ans: -

* P(x>0) = 0.6, implies there is a 60% chance that the venture would yield profits or greater than expected returns. P (Incurring losses) is only 0.2.
* So, the venture is likely to be successful.

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

Ans: -

* Weighted average = x\*P(x) = 800. This means the average expected earnings over a long period of time would be 800(including all losses and gains over the period of time)

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

* Ans:- The good measure of the risk involved in a venture of this kind depends on the variability in the distribution. Higher Variance means more chances of risk Var(X) = E(X^2)-(E(X))^2 = 2800000 – 800^2 = 2160000