**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% |

**Refer Jupyter Note Book**



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.

**Lets Consider Q1 = 5; Q3 = 12; So, IQR = Q3-Q1 = 12-5 = 7**

**The value is important to summarize the spread of data, to identify outliers and to understand the distribution of data with respect to its central part i.e 50% of data**

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

**It is also known as Positively Skewed. In this type of distribution the tail extends towards the right and the more data is on the left side.**

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?

**The datapoint with the value 25 to 2.5 would lead to a change in boxplot. The whisker on the right side would be shorter. Also, the distribution of data would be more accurate without the influence of the outlier.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**The Mode would be (20 or 22) for Y-axis & between 4 to 8 with respect to values in X-axis according to the graph.**

1. Comment on the skewness of the dataset.

**It is a positive skewed dataset. It has a long tail at the right. The mean of the dataset would be greater than the median. The dataset is Asymmetric.**

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.

**The Histogram & Boxplot both highlights the outliers and also both concludes that the distribution is right-skewed or positively skewed.**

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.

**If 1 in 200 long-distance telephone calls are getting misdirected.**

**probability of call misdirecting   = 1/200**

**Probability of call not Misdirecting = 1-1/200 = 199/200**

**The probability for at least one in five attempted telephone calls reaches the wrong number**

**Number of Calls = 5**

**n = 5**

**p = 1/200**

**q = 199/200**

**P(x) = at least one in five attempted telephone calls reaches the wrong number**

**P(x) = ⁿCₓ pˣ qⁿ⁻ˣ**

**P(x) = (nCx) (p^x) (q^n-x)**

**P(1) = (5C1) (1/200)^1 (199/200)^5-1**

**P(1) = 0.0245037**

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?

**Here it would be 2000 ( w.r.t nearest 1000)because it has a probability of 0.3, which is the highest among the all.**

1. Is the venture likely to be successful? Explain

**We must calculate average,**

**Avg = sum[x\*P(x)] = (-2000\*0.1) + (-1000\*0.1) + (0\*0.2) + (1000\*0.2) + (2000\*0.3) + (3000\*0.1) = -200 – 100 + 0 + 200 + 600 + 300 = 800**

**Hence the average is $800 in positive, we can likely consider it to be successful.**

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

**The long-term average earning are $800 per venture. Which means, on average, per venture or each venture of this kind is expected to earn $800.**

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

**Var (X) = E(X^2) – (E(X))^2**

**= 2800000 – 800^2**

**= 2160000**