**Set-1**

**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:- Refer Jupiter Notebook File.**



**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.**
2. **What can we say about the skewness of this dataset?**
3. **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:-**

1. **Assume,**

**Lower Quartile (LQ) = 5,**

**Upper Quartile (UQ) = 12**

**Inter Quartile Range (IQR) = UQ – LQ = 12-5 = 7**

**Inter quartile range is the difference between upper quartile and lower quartile.**

1. **Here right whisker is longer than the left whisker, so the data is positively skewed.**
2. **Then that data point will be fall in left whisker and value of median will be change.**



**Answer the following three questions based on the histogram above.**

1. **Where would the mode of this dataset lie?**
2. **Comment on the skewness of the dataset.**
3. **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:**-

1. In between 4 and 8.
2. Data is positively skewed.
3. In both the graphs value of data point 25 is seen far away from other values so we can call it as an outlier. In boxplot right whisker is longer than the left so it is positively skewed, and in histogram also data is positively skewed.
4. **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:-**

1 in 200 long-distance telephone calls is misdirected so the probability of call misdirecting.  p = 1/200.

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

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

At least one in five attempted telephone calls reaches the wrong number

P(X) = 1 - none of the call reaches the wrong number

P(x) = 1 - P(0)

P(x) = 1 - ⁵C₀(1/200)⁰(199/200)⁵⁻⁰

P(x) = 1 - (199/200)⁵

P(x) = 0.02475

Probability that at least one in five attempted telephone calls reaches the wrong number = 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?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

**Ans:-**

|  |  |  |
| --- | --- | --- |
| **X** | **P(X)** | **E(X) = X \* P(X)** |
| **-2000** | **0.1** | **-200** |
| **-1000** | **0.1** | **-100** |
| **0** | **0.2** | **0** |
| **1000** | **0.2** | **200** |
| **2000** | **0.3** | **600** |
| **3000** | **0.1** | **300** |

**Expected Value = E(X) = ∑ ( X \* P(X)) = 800.**

1. **Most likely monetary outcome of the business venture is $2000 as it has maximum probability = 0.3.**
2. **Probability P(X>0) = 0.60, there are 60% chances of profit.**

**Probability P(X<0) = 0.20, there are 20% chances of loss.**

**Here percentage of profit is greater than percentage of loss, so we can say that the venture is likely to be successful.**

1. **Long term average earning of business ventures = $800.**
2. **Standard Deviation (σ).**

**Mean = (-2000-1000+0+1000+2000+3000)/6**

**Mean = 500.**

**n-1 = 6-1 = 5**

**Std. dev = √ [(-2000-500)2+(-1000-500)2+(500-0)2+(1000-500)2+(2000-500)2+(3000-500)2]/5**

**Std.dev = 1870.8**