**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% |

**Mean : 0.33271333333**

**Standard deviation: 0.16370812590**

**Variance : 0.026800350488888**

**Outliers : [0.9136]**



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 : Q3-Q1

12.5 – 5

7.5

**IQR contains 50% of datapoints**

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

**Above given data is Right Skewed.**

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?

**Nothing will be affected as new datapoint won’t change anything.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Mode lies between 4-8.**

1. Comment on the skewness of the dataset.

**Right Skewed.**

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: In histogram we get central tendency as mode with peak values where as in box plot we get median value as central tendency.**

**In boxplot with IQR and Whiskers length we can get idea of data variability.**

**Whereas in histogram the width of bins will help us to asses the data variability.**

**In histogram we can identify outliers with lsolated bar from main distribution whereas in box plot we can easily identify outliers.**

**Skewness can be visually seen with the asymmetry of the histogram where in boxplot we get a hint of asymmetry if one whisker is significantly longer than other.**

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 =**

**Probability of success : p = 1/200**

**Probability of failure : q = 1-p = 199/200**

**Using the binomial formula**

**P(x) = x qn-x**

**Where n is number of trails.**

**X is number of successful trails**

**P is probability of success**

**Q is probability of failure**

**P(0) =**

**= 0.02475**

**The probability of at least one call being misdirected is 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?

**Highest probability is 0.3 , so 2000 is most likely monetary outcome.**

1. Is the venture likely to be successful? Explain

**Positive outcomes = 1000, 2000 and 3000.**

**Sum of probabilities = 0.2+0.3+0.1 = 0.6.**

**The cumulative probability of positive outcomes is greater than 0.5 hence the venture likely to be successful.**

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

**Long term average earnings is 800.**

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

Mean is 800

Calculate squared difference and multiply by the probabilities for each outcome

**(-2000-800)2  \* 0.1 =784000**

**(-1000-800)2 \* 0.1 = 32400**

**(0-800)2 \* 0.2=12800**

**(1000-800)2 \* 0.2 = 8000**

**(2000-800)2 \* 0.3 = 432000**

**(3000-800)2 \* 0.1 = 484000**

**Sum of squared difference = 1868400**

**Calculate the square root of the sum divided by the total number of observations:**

**Standard deviation: 558.03**

**This measure of risk indicates the variability or spread of monetary outcomes around mean.**