**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 :-**

1. Outlier is Morgen Stanley 91.36%
2. µ = 33.27133
3. σ = 16.9454
4. σ2 = 287.1466



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. Interquartile range: Q3 -Q1 = 12-5 = 7

This value implies median value of above boxplot.

1. Dataset is positively skewed.
2. In that case there would be no outliers and so, data will be normally distributed.



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. Mode of this dataset lie between 4 to 8.
2. Right Skewed.
3. This Histogram and boxplot in Q2 both are right skewed and both have outliers. The median can be easily visualized in the box plot where as in histogram mode is more visible.
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 :-**

Let X be the event as,

X :- The call is misdirected.

P(X) :- 1/200

Hence,

Y :- The call is getting connected correctly.

P(Y) :- 199/200

Probability that at least 1 call in 5 attempts reaches the wrong number

= 1-probability that no attempted call reaches the wrong number

= 1-(199/200)5

= 0.025

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 :-**

1. From the given data we get the most likely monetary outcome of the business venture is 2000 because it’s probability is high.
2. The venture is likely to be successful because probability is increasing.

The probability is that the venture will make more than 0 or a profit is

P(x>0)+p(x>1000)+p(x>2000)+p(x=3000)

= 0.2+0.2+0.3 = 0.7

1. Long Term Average earing of business venture of given kind

=

*(-*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

1. The good measure of the risk involved in a venture of this kind is dependent on variability in the given distribution. Greater is the variance implies more chance of risk. **A good measure to evaluate the risk would be variance and standard deviation of the variable x**

σ=sqrt{ ∑i(xi−μ)^2⋅P(xi) }

=sqrt{ ( (-2000-800)^2\*0.1)+((-1000-800)^2\*0.1 )+((0-800)^2\*0.2)+((1000-800)^2\*0.2)+((2000-800)^2\*0.3)+((3000-800)^2\*0.1) }

= 1190.24

The standard deviation, a measure of risk, is approximately $1,190.24.