**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% |

* Answer: - Outlier: - 91.36

Mean: - 33.2713

Standard Deviation: - 16.945400921222028

Variation: - 287.1466123809524



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.

* Answer: -

Approximately First Quantile Range i.e. Q1 = 5

Third Quantile Range i.e. Q3 = 12

Inter Quartile Range (IQR) = Q3 – Q1

= 12 – 5

IQR = 7

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

* Answer: -
* Positively Skewed
* The data is Right Skewed and the median is towards left side so it is not a normal distribution.

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?

* Answer: -
* In that case there would be no outliers on the given dataset
* Because of the outlier the data had positive skewness it will reduce and the data will normal distributed.
* Also the values of mean and median would slightly changes.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* Answer: -

The mode of the dataset is lies between 5 to 10 and approximately between 4 to 8

1. Comment on the skewness of the dataset.

* Answer: -

The Dataset is Right Skewed.

Also the data is Positively 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.

* Answer: -
* Boxplot gives us the Median while Histogram gives us the Mode.
* Histogram provides us the frequency distribution so we can see how many times each data point is occurring.
* Boxplot provides the quantile distribution i.e. 50% data lies between 5 and 12.
* Also boxplot provides us the whisker length to identify the outliers, and from histogram we can only guess looking at the gap that 25 may be the 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.)

* Answer: -

Probability of call getting misdirected = (1/200)

Hence Probability of call not getting the misdirected = 1 – (1/200)

= 199/200

Number of phone calls attempted = 5

Therefore, probability that at least one in 5 attempted call reaches the wrong number is:

= 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?

* Answer: -

Max. P = 0.3 for P (2000)

So most likely outcome is 2000

1. Is the venture likely to be successful? Explain

* Answer: -

Venture is successful if X is +ve

Hence, if X is 1000, 2000 or 3000

Probability is 0.2 + 0.3 + 0.1 = 0.6

As 0.6 > 0.5 **hence Venture is likely to be successful**

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

* Answer: -

The long term average is Expected value = Sum (X \* (PX))

= 800$

Which means on an average the return will be +800$

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

* Answer: -

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 (Quite High)

SD = √Var  ≈ **$ 1470**

As variability is quite high hence risk is high