**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 Jupyter notebook for more details. Following are outliers found in boxplot:

Morgan Stanley 91.36%,

Mean = 33.271333

Standard deviation = 16.945401

Variance = 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.

Ans: From above plot we can see that Q1(1st Quartile R) = 5, Q3(3rd Quartile R) = 12, and Q2-Median (2nd Quartile R) = 7

IQR (Inter-Quartile Range) = Q3 - Q1 = 12 - 5 = 7

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

Ans: It is Right-skewed-Mean>Median>Mode (Positive Skewness) median towards left side which means it not normally distributed i.e. skewness > 0

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?

**Ans:** The present outliers of data have positive skewness, if data point is 2.5 then in that case there won’t be **NO Outliers** on given dataset which will reduce & data will get normally distributed.

1. 

Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans:** The mode of this dataset lies between 5 to 10.

1. Comment on the skewness of the dataset.

**Ans:** It is Right-skewed-Mean>Median>Mode (Positive Skewness) median towards left side which means it not normally distributed i.e. skewness > 0

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:** Both are Right-skewed-Mean>Median>Mode (Positive Skewness) as well as both have outliers.

In box plot, median is easily visualized where as In histogram, mode is more visible.

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

Probability of call misdirecting (p) = 1/200

Probability of call not misdirecting (1-p) = 1-1/200 = 199/200

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

No. of Calls = 5,

Where; n = 5, p = 1/200, q = 199/200, x = 1

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) # nCr = n! / r! \* (n - r)!

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 |

By using this formula E(X) = x\*P(x) and E(X^2) =x^2\*P(x) below are results;

|  |  |  |  |
| --- | --- | --- | --- |
| x | P(x) | E(X) | E(X^2) |
| -2,000 | 0.1 | -200 | 400000 |
| -1,000 | 0.1 | -100 | 100000 |
| 0 | 0.2 | 0 | 0 |
| 1000 | 0.2 | 200 | 200000 |
| 2000 | 0.3 | 600 | 1200000 |
| 3000 | 0.1 | 300 | 900000 |
| Total | | 800$ | 2800000$ |

1. What is the most likely monetary outcome of the business venture?

**Ans:** The most likely monetary outcome of the business venture is 2000$

As probability for 2000$ is 0.3 which is maximum amongst all.

1. Is the venture likely to be successful? Explain

**Ans:** Yes.

The probability that venture can make profit;

P = p(x>0) + p(x>1000) + p(x>2000) +p(x=3000) = 0.2 + 0.2 + 0.3 + 0.1 = 0.8 (i.e. 80%)

Which indicates that there is 80% chances for this venture to make a good profit.

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

**Ans:** By adding E(X) we will get long term on an average i.e. 800$

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

**Ans:** The good measure of the risk involved in a venture of this kind will depend on variability in Distribution i.e. Variance. Higher the Variance, more will be the risk (prediction quality is poor).

**Var (X)** = E(X^2) –(E(X)) ^2

= 2800000 – 800 ^2

**= 2160000**