**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:

Mean() = (24.23 + 25.53 + 25.41 + 24.14 + 29.62 + 28.25 + 25.81 + 24.39 + 40.26 + 32.95 + 91.36 + 25.99 + 39.42 + 26.71 + 35.00) / 15

= 33.271333

Varience() = 287.146612 (by Using Varience formula)

Standard Deviation() = 16.945401

**(Refer IPYNB)**



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. By seeing the boxplot we can say that approximately

Lower quartile = 5 &

Upper quartile = 12(approx.)

We know that formula for IQR = Upper quartile – Lower quartile

IQR = 7 (approx.)

Lower Extreme = Lower quartile – (1.5)\*IQR

Upper Extreme = Upper quartile + (1.5)\*IQR

Lower Extreme = 5 – (1.5 \* 7) = 5 – 10.5

= -5.5

Upper Extreme = 12 + (1.5 \* 7) = 12 + 10.5

=22.5

Hence inter-quartile range of this dataset is (-5.5 to 22.5)

1. Since the most of the data lies on after 50% we can say that mean is more than median, also we know that if mean is more than median then the it is Right Skewed
2. New box plot will not be affected because there is no outliers in the data hence we can say that data is 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. The mode of this dataset lies in (approx.) 4,5,6,7.

Because if we observe the histogram there is more repitation in neighbours of 5th datapoint

1. By seeing the Histogram, there is more data on right compared to left, and also by observing we can say that mean is more than median hence it is right skewed
2. If the above histogram and box-plot are plotted for the same dataset we can say that box plot will give us the outliers (which was problem for us), and the histogram will tell us how the data are distributed to know the data better
3. 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 getting misdirected.

probability of call misdirecting = 1/200

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

The probability for at least one in five attempted telephone calls reaches the wrong number Number of Calls = n = 5

P(success) = 1/200

Q(fail) = 199/200

We know that binomial distribution,

P(x) = at least one in five attempted telephone calls reaches the wrong number

P(x)=(nCx) (p^x) (q^n-x)

Where, nCr = n! / r! \* (n - r)!

P(1) = (5C1) (1/200)^1 (199/200)^5-1 P(1) = 0.02475(approx.)

Hence, the probability that at least one in five attempted telephone calls reaches the wrong number is 0.02475(approx.)

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. The most likely monetary outcome is the one with the highest probability. In this case, the outcome with a probability of 0.3 (or 30%) is a return of $2000
2. The venture is likely to be successful because there is a 80% chance (the sum of probabilities for outcomes 0, $1000, $2000, and $3000) of making a profit (positive return),

which indicates that more often than not, the venture will yield a positive result.

1. To find the long-term average earnings, multiply each outcome by its probability, then sum the results:

Expected Value=(−2000×0.1)+(−1000×0.1)+(0×0.2)+(1000×0.2)+(2000×0.3)+(3000×0.1)

Expected Value=−200−100+0+200+600+300=800

So, the long-term average earning for business ventures of this kind is $800.

1. To measure the risk involved in a venture of this kind, you can calculate the standard deviation. The standard deviation measures the dispersion or spread of the outcomes around the mean. A higher standard deviation indicates greater risk.

To calculate the standard deviation, you'll need to use the following formula:

Standard Deviation (σ) = √[Σ[(x - E)^2 \* P(x)]]

Where:

x is each possible outcome.

E is the expected value (calculated in part iii).

P(x) is the probability of each outcome.

Let's calculate it step by step:

σ = √[(-2,000 - 800)^2 \* 0.1 + (-1,000 - 800)^2 \* 0.1 + (0 - 800)^2 \* 0.2 + (1,000 - 800)^2 \* 0.2 + (2,000 - 800)^2 \* 0.3 + (3,000 - 800)^2 \* 0.1]

σ ≈ √[784,000 + 324,000 + 128,000 + 8,000 + 864,000 + 484,000]

σ ≈ √(2,588,000)

σ ≈ 1,608.88 (rounded to the nearest dollar)

So, the standard deviation is approximately $1,609. This is a measure of the risk involved in the venture. The higher the standard deviation, the greater the risk. In this case, there is a significant spread in potential outcomes, indicating a relatively high level of risk.