**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% |

In the given above data by seeing Boxplot we can say that there is only one outlier that is 91.36

Mean value () = 33.27133333333333

Standard Deviation = 16.370812590976932

Variance = 268.00350488888887



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?
4. Ans: Q3-Q1 is the inter quartile range which says that there will be overall 4 whiskers and each whisker consists of 25% of the data and from this diagram Q1 is at point 5 and Q3 is at 12. So overall data is present in between 0 to 19.

12 - 5 = 7

Any value beyond (1.5) IQR is an outlier

i.e., Q1 – (1.5) IQR will be considered as an outlier and at the same time Q3 + (1.5) IQR will be considered as and outlier.

1. By this given dataset we can clearly say the distribution of the data is Right Skewed data and

yes, there is an outlier in the given dataset and also, we can say that most of the data is

present in the 3rd and 4th whiskers. That’s the reason we can see the distribution is right

skewed.

1. If data point 25 is actually 2.5 then there would not be any outlier and that’s not sure because any other data point end to the whisker might affect the dispersion of the data.

We may get the boxplot like the same right skewed data and that’s only because we are considering only one data point in the outlier as the inlier.



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.
4. From the histogram we can say that mode of the data set lies between 5 and 8 because we can say that the peaked ness of the data lies between the point 5 and 8 that’s the reason we will be considered as mode.
5. From the histogram we can clearly say that the distribution of the data is right skewed data. That’s because most of the data lies right side of the median and mean that’s the reason, we can call it as Right skewed data distribution.
6. From the both Boxplot and histogram we can say that it’s same dataset graphs because both are right skewed and there is only one outlier and that lies on the point 25 in both graphs so these both articulates that both are right skewed data and both has only one outlier and also most of the data lies between the point 5 and 8. Mode also lies between 5 and 8 and also Mean and Median also lies as same as in the graphs each other.
7. 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.)

Probability of at least one in the five attempts reaches the wrong number call.

So, we can use complement probability.

We will be finding out the probability that no call among the five are misdirected and then we will have to subtract the probability from 1.

We are given with 200 long distance call and those are misdirected.

So, from the probability that a single telephone call is not misdirected.

P (Not Misdirected) = 1 – P(Misdirected) = 1 – (1 / 200) = 199/200

As given if we consider five attempts of the calls then we can say that these attempts to be independent, the probability of no call of the five calls are misdirected is the multiplication of the probabilities that each individual call is not misdirected.

P (All the five not misdirected) = (199 / 200)^5

P (at least one misdirected) = 1 – (199 – 200)^5

= 0.024 approximately

So, the probability is 0.024 or we can also say 2.4%

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
5. To know the most likely monetary outcome of the business venture, first we need to know the highest value in the P(x) or we can also say that mode of P(x), and from the given data we can see mode is 0.3 that is the x is 2000. So that we can conclude that most likely monetary outcome of the business would be 2000.
6. To determine if the venture is likely to be successful, we need to consider the expected value (also known as the mean) of the monetary outcomes.

Expected Value (EV) = Σ [x \* P(x)]

EV = (-2,000 \* 0.1) + (-1,000 \* 0.1) + (0 \* 0.2) + (1,000 \* 0.2) + (2,000 \* 0.3) + (3,000 \* 0.1)

EV = (-200) + (-100) + (0) + (200) + (600) + (300) = 800

The expected value is $800. Since the expected value is positive, it suggests that, on average, the venture is likely to be successful over the long term. However, success also depends on the level of risk associated with the venture.

1. The long-term average earnings of business ventures of this kind can be calculated as the expected value i.e., mean of the probability distribution. To find the expected value (E(x)), you multiply each possible outcome by its probability and then sum them up:

E(x) = (-2,000 \* 0.1) + (-1,000 \* 0.1) + (0 \* 0.2) + (1,000 \* 0.2) + (2,000 \* 0.3) + (3,000 \* 0.1)

E(x) = -200 - 100 + 0 + 200 + 600 + 300 = 800

So, the long-term average earnings of business ventures of this kind are 800.

To measure the risk involved in a venture of this kind, we can calculate the standard deviation (σ) of the distribution. The standard deviation provides a measure of how spread out or variable the outcomes are from the expected value.

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

Where μ is the expected value (calculated in part ii), x is each possible outcome, and P(x) is the probability of that outcome.

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 ≈ √[(256000 \* 0.1) + (32400 \* 0.1) + (256000 \* 0.2) + (32400 \* 0.2) + (256000 \* 0.3) + (484000 \* 0.1)]

σ ≈ √[25600 + 3240 + 51200 + 6480 + 76800 + 48400]

σ ≈ √[183,920]

σ ≈ 429.09

So, the standard deviation is approximately $429.09. This measures the level of risk in the venture. A higher standard deviation indicates higher risk, and in this case, it suggests that there is some level of variability in the returns, which means there is risk associated with the venture.