**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-** Variance σ2 = 268.0035 or 2.87%

Standard Deviation σ = 16.370813 or 0.16%

Count n = 15

Mean μ = 33.271333

As shown in graph, the outlier is hold by Morgan stanley Measure of 91.36%



**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 –** The inter-quartile range will be Q3-Q1= 13-5= 8 (approx)

The value implies that the majority of the data is concentrated towards the higher end i.e right hand of the scale, while a smaller proportion of the data is spread out towards the lower end.

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

**Ans –** The skewness of this dataset is right-skewed or positively skewed 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?**

**Ans -** If the data point with the value 25 is actually 2.5, it means that there is an error in the data. This data point is an outlier, and it will affect the box-and-whisker plot by shifting the whisker on the right-hand side and changing the scale of the plot.

The box-and-whisker plot would still show a Right-skewed distribution, but the scale and position of the plot would be adjusted to accommodate the corrected data point.



**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 would lie in the left of the graph.

1. **Comment on the skewness of the dataset.**

**Ans -** The skewness of this dataset is right-skewed or positively skewed distribution.

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-** Together, these graphs complement each other by providing a more complete picture of the dataset. The histogram gives an overall picture of the frequency distribution, while the box plot provides a more condensed and focused view of the key measures that describe the distribution & the outliers.

**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 -** This problem can be solved using the binomial distribution, which is appropriate for counting the number of successes in a fixed number of independent trial.

P(at least one call is misdirected) = 1 - P(all five calls are correctly directed)

= 1 - (1-p)^5

= 1 - (199/200)^5

= 0.0247

Therefore, the probability that at least one in five attempted telephone calls reaches the wrong number is approximately 0.0247, or about 2.47%.

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?**

**Ans-** The most likely monetary outcome of the business venture is the value with the highest probability, which is $2,000 with a probability of 0.3.

1. **Is the venture likely to be successful? Explain**

**Ans-** If success is defined as a positive return, then the venture is likely to be successful since the sum of the probabilities for positive returns (i.e., returns greater than zero) is 0.6, which is greater than the sum of the probabilities for negative returns (i.e., returns less than or equal to zero) which is 0.4.

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

**Ans -** The long-term average earning of business ventures of this kind can be calculated as the expected value of the returns. That is:

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) = $800

The long-term average earning of business ventures of this kind is $800.

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

**Ans -** One measure of the risk involved in a venture of this kind is the standard deviation of the returns.

The formula for the standard deviation of a discrete probability distribution is:

σ = sqrt[Σ(x - μ)^2 P(x)]

σ = sqrt[(2000 - 900)^2(0.3) + (1000 - 900)^2(0.2) + (0 - 900)^2(0.2) + (-1000 - 900)^2(0.1) + (-2000 - 900)^2(0.1) + (-3000 - 900)^2(0.1)]

σ = $1,469.69

the standard deviation of the returns is $1,469.69, which is a measure of the risk involved in a venture of this kind.