**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

=

Variance = ((24.23- )^2+(25.53- )^2+(25.41- )^2+(24.14- )^2+(29.62- )^2+(28.25- )^2

(25.81- )^2+(24.39- )^2+(40.26- )^2+(32.95- )^2+(91.36- )^2+(25.99- )^2

(39.42- )^2+(26.71- )^2+(35.00- )

=

Standard deviation = sqrt(variance) = sqrt( )

So, the mean,variance, and standard deviation of the data are µ = , σ^2 = ,

σ = respectively.

Note : In this data the value of Morgan Stanley (91.36%) is an outlier, as it is significantly different from

The rest of the data. This can affect the mean and standard deviation of the data, making them

Less representative of the actual distribution of the data.



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) Inter-quartile = Q3-Q1 = 12 – 5 = 7

The inter-quartile range represents the range of values that encompasses more on the left

Of the data.

(2) based on the shape of the box-plot, it can be seen that the data is skewed to the right, as

The right tail extends further away from the median than the left tail. So, it is positively

Skewwd.This indicates that There are some extreme values on the higher end of the

range of data.

(3) If the value 2.5 was actually 25, then this value would be within the range of the middle

50% of the data and would not affect the skewness of the dataset. However , it would shift

The median slightly to the right and increase the inter-quartile range slightly.

Top of Form



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 a dataset is the value that appears most frequently. In a histogram, the mode

Can be determined by identifying the peak or tallest bar, which represents the most

Common values. Here, Mode = 21 (approximately) because the most frequency value

Occured above the value 20.

2) The skewness for the above histogram is positively skewed.

3) A histogram and a box-plot can complement each other in providing information about a

Data set by showing different aspects of the distribution of the data. A histogram provides

A visual representation of the distribution of the data and can show the frequency of the

Each value in the dataset. A box plot on the other hand, provides a summary of the

Distribution of the data, including the median, quartiles and outliers. Together, a histogram

And a box plot can give a comprehensive picture of the distribution of the data and allow

For a more informed analysis.

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) We can use the binomial distribution to calculate the probability that at least one in five

Attempted telephone calls reaches the wrong number. In binomial distribution models the

Number of successes in a fixed number of independent trials, where each trial has the same

Probability of success.

Let X be the number of misdirected calls in five attempts. Then X follows a binomial distribution

With parameters n = 5 and p = 1/200.

To find the probability that at least one in five attempted telephone calls reaches the wrong

Number, we can calculate the probability that X is greater than or equal to 1, which is given by :

P(X>=1) = 1 – P(X = 0)

The probability mass function of the binomial distribution can be used to calculate

P(X=0) = (n choose 0)\*p^0\*(1-p)^(n-0)

= (5 choose 0) \* (1/200)^0 \* (199/200)^5

= 0.9752

Finally, we can calculate the probability that at least one in five attempted telephone calls reaches the

Wrong number :

P(X >= 1) = 1 – 0.9752 = 0.0248

So, the probability that at least one in five attempted telephone calls reaches the wrong number is

Approximately 0.0248 or 2.48% .

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 of the business venture is $2,000, as it

Has the highest probability of 0.3.

(2) The success of a business venture is subjective and depends on the

Financial goals and expectations of the business. In this case, if the

Expected return is greater than $2000, the venture may be considered

Successful. If the expected return is less than $2000, the venture may be

Considered unsuccessful.

(3) The long – term average earning of business ventures of this kind can be

Calculated using the expected value, which is the sum of the product of

Each outcome and its probability. The expected value of the business

Venture can be calculated as follows :

(-2000)\*(0.1)+(-1000)\*(0.1)+(0)\*(0.2)+(1000)\*(0.2)+(2000)\*(0.3)+(3000)\*(0.1)=$800

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

(4) A good measure of the risk involved in a venture of this kind is the

Standard deviation, which measures the spread of the data around the

Mean. The standard deviation of the business venture can be calculated

As follows:

Variance = (x1^2\*P(x1)+x2^2\*P(x2)+…+xn^2\*P(xn)) – (mean)^2

Where x1,x2,…,xn are the outcomes and P(x1),P(x2),…,P(xn) are the probabilities.

Variance = (-2000^2\*0.1+ -1000^2\*0.1+ 0^2\*0.2+1000^2\*0.2+2000^2\*0.3+3000^2\*0.1) – (800)^2 = 2160000

The standard deviation is the square root of the variance :

Standard deviation = sqrt(Variance) = sqrt(2160000) = 1469.6938

So, the standard deviation of the business venture is $1500, which indicates that the risk

Involved in a venture of this kind is relatively high.