**Topics: Descriptive Statistics and Probability**

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% |

24.14%

24.23%

24.39%

25.41%

25.53%

25.81%

25.99%

26.71%

28.25%

29.62%

32.95%

35.00%

39.42%

40.26%

91.36%

N=number of data point

Lower Quartile (**Q1**) = (N+1) \* 1 / 4

(15+1)\*1/4=4thposition = 25.41%

Upper Quartile (**Q3** )= (N+1) \* 3 / 4

(15+1)\*3/4=12thposition= 35.00%

**IQR**= Q3-Q1

= 35.00% - 25.41%

=9.59%

U.E=Q3+1.5\*IQR

=49.385% Any value above than this are considered outliers

Morgan Stanley 91.36% is outlier

L.E =Q1-1.5\*IQR

=11.025% Any value below than this are considered outliers

population mean = = 33.27

Standard deviation = = 16.37

population variance = = 268.00



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.

Approximately (First Quantile Range) **Q1** = 5 ,(Third Quantile Range) **Q3** = 12

, **Median**(Q2)= 7

**IQR=** Q3-Q1= 12-5 **=**7

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

Firstly, It is not Normally Distributed

right wisker lenth is greater than left wisker length

so we can say it is showing the Right Skewness in this dataset

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?

In that case there would be no outliers on the given dataset because of the outlier the data had positive skewness it will reduce and the data will normal distributed.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

The mode of this data set lie in between 5 to 10 and approximately between 4 to 8

1. Comment on the skewness of the dataset.

Right-Skewed. Mean>Median>Mode

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.

They both are right-skewed and both have outliers the median can be easily visualized in box plot 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: X=1 in 200 long-distance telephone calls are getting misdirected.

then probability of the event X is

P(X)= 1/200

X bar =Probability of call not Misdirecting = 1-1/200

P (X bar) = 199/200

Therefore,

Probability that at least one in 5 attempted call reaches the wrong number

= 1 - Probability that no attempted call reaches the wrong number

= 1 – P (X bar)

= 1 – (199/200) \* (199/200) \* (199/200) \* (199/200) \* (199/200)

= 1 - (199/200) ^5

= 0.025

Probability that at least one in 5 attempted call reaches the wrong number = 0.025

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. 2000

1. Is the venture likely to be successful? Explain

Ans. Yes, the probability that the venture will make more than 0 or a profit

p(x>0) +p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 this states that there is

a good 80% chances for this venture to be making a profit

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

Ans. The long-term average is Expected value = Sum (X \* P(X)) = 800$ which means on an

average the returns will be + 800$

E(X) =Sum X.\*P(X)

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

= 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 depends on the Variability

in the distribution. Higher Variance means more chances of risk

E(X^2) = X^2\*P(X)

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

= 2800000

Var (X) = E(X^2) –(E(X))^2

= 2800000 – 800^2

= 2160000