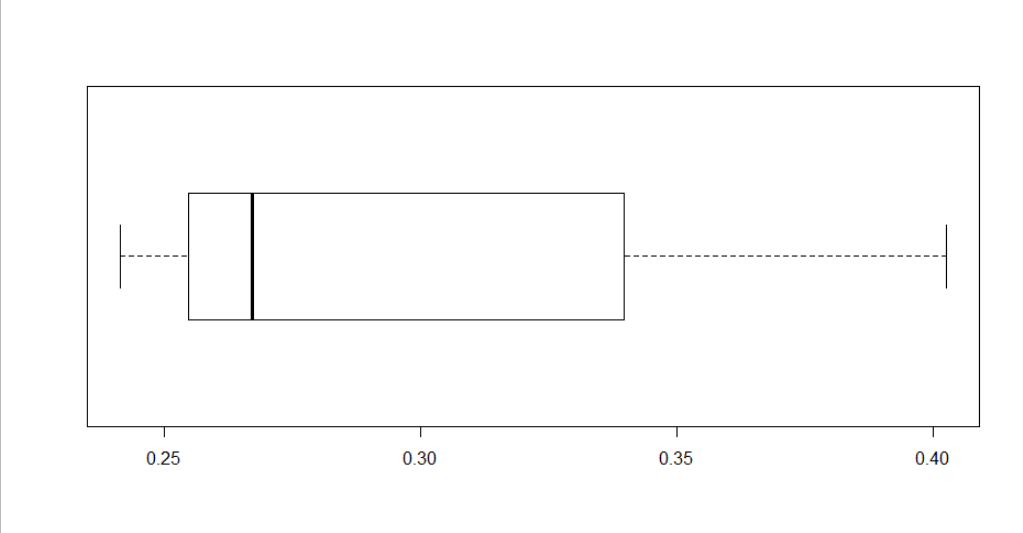
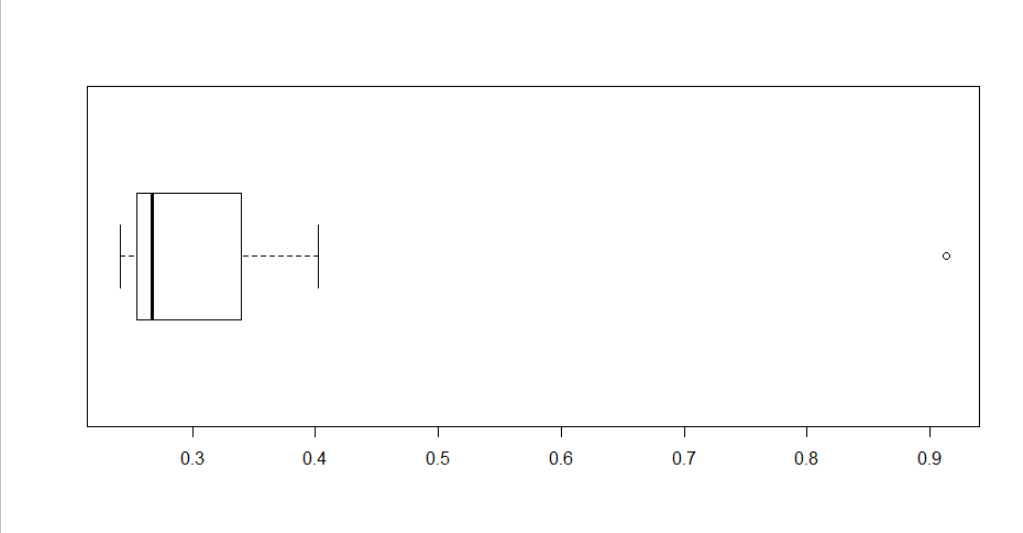
**Topics: Descriptive Statistics and Probability**

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% |

Explanation:

**Excluding Outliers** **Including Outliers**

Calculation: firstly we convert this data in the decimal points

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Minimum Vale** | **1 quartile(Q1)** | **Median** | **Mean()** | **3rd quartile(Q3)** |
| 0.2414 | 0.2547 | 0.2671 | 0.3327 | 0.3423 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Standard Deviation)** | **Variance** | **Maximum Value** | **IQR=Q3-Q1** |  |
| 0.169454 | 0.02871466 | 0.9136 | .0876 |  |

**Upper Major Outlier=0.6051**

**Upper Minor Outlier=0.4737**

* Main purpose of plotting Boxplot is identify Outliers, data spread or shape of data like Skewness and Kurtosis and also useful for calculate median of the data.
* Median of the data is 0.2671
* Skewness: Positive Skewed , data concentrate on the left side
* Kurtosis: Positive kurtosis
* As one of outliers (Major outliers) lies in the data.



Answer the following three questions based on the box-plot above.

**Explanation: Lower Quartile(Q1)=5,Upper Quartile(Q3)=12**

1. What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

**Explanation:** Inter Quartile Range (IQR)=Q3-Q1=12-5=7

7 is the IQR of the Data Set which means the box contain 50% of the data points.

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

**Explanation:** Positive Skewness ,which means data concentrate on left side and long tail on right side .

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?

**Explanation:** If we consider that the value 25 is actually 2.5 then the data points that will be in the data set are 0,0.5,1.0,1.5,2,2,2.5 then upper quartile =1.2 , lower quartile= 0.5

Then inter quartile range (IQR)=Q3-Q1=1.2-0.5=0.7



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Explanation: The Mode of the dataset is 6.

1. Comment on the skewness of the dataset.

Explanation: Positive Skewness or Right Skewness, which means that data spread on the right side and data concentrµate on left side.

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.

Explanation: histogram and boxplot graph complement each other in this case like both graph shows positive skewness, Most of the data points contain in (12-5) range.

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.)

**Explanation:**

Let us define an event

A: The call is misdirected

then probability of the event A is

P(A)=1/200

Therefore

P(=1-P(A)=1-1/200=199/200

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

=5/200=0.025

“Probability that at least one in 5 attempted call reaches the wrong number is **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?

The most likely monetary outcome of the business venture is 3000

1. Is the venture likely to be successful? Explain

**Explanation:**

Yes! Because here the total loss probability is 0.1 + 0.1 = 0.2 and 0.2 is the probability in which there is no loss nor profit and the total profit probability is 0.2 + 0.3 + 0.1 = 0.6 so we can say that total profit probability > total loss probability

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

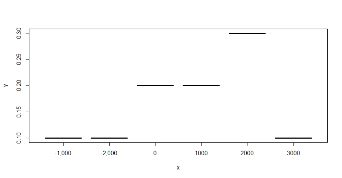
we can calculate the long-term average earning of business ventures of this kind so here we calculate the average so

µ = x1.p(x1)+x2.p(x2)+x3.p(x3)+x4.p(x4)+x5.p(x5)+x6.p(x6)

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

µ = 800

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



**Explanation:** By this graph we can see that from 2000 probability of 30% and it is decreased by 10% of the 3000 which shows that volatility of the expected data.

We have consider how far each value is from our mean data and need to see how it vary from average. We can calculate Variance and Standard deviation to assume data distance from our average profit or loss.

Average=500 , Median=500, Variance= 2916667 Standard Deviation= 1 1707.825

By the variance we can find our data points, distance from average and by the median can find middle value of the data.