**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% |

company<-read.csv(file.choose())

View(company)

attach(company)

mean(Measure.X) #mean of measure X = 33.27%

sd(Measure.X) #standard deviation of measure X = 16.95%

var(Measure.X) #variancee of measure X = 287.15%

library(ggplot2)

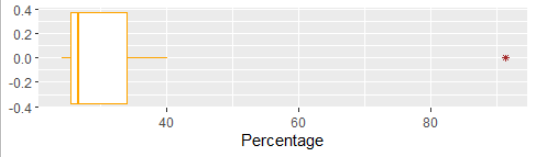
#BOX PLOT

g2 <- ggplot(company, aes(Percentage))+ geom\_boxplot(col = "orange",outlier.colour="brown", outlier.shape=8,outlier.size=1)

plot(g2)

g2 + ggtitle("BOX PLOT OF %X VALUES")

# ONE OUTLIER, 91.36% FOR Morgan Stanley COMPANY

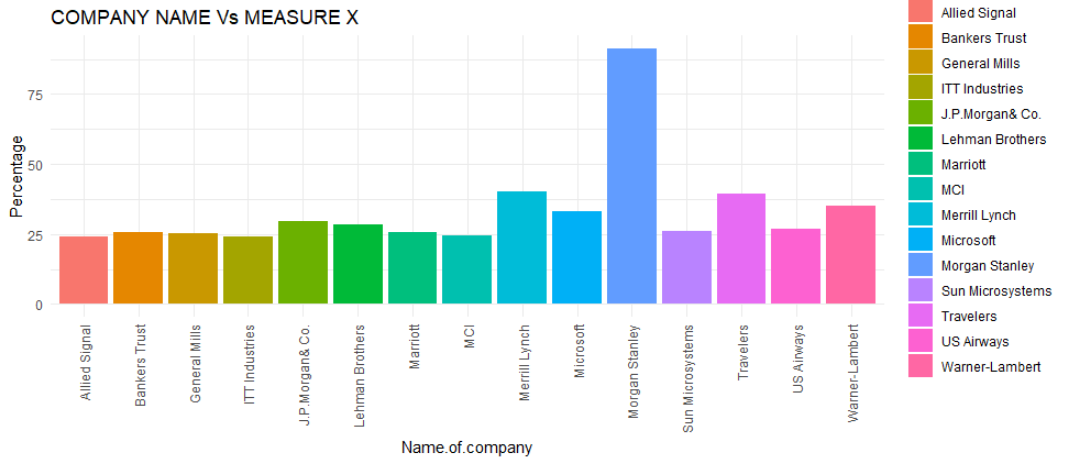


#BAR CHART

g1 <- ggplot(company, aes(x=Name.of.company, y=Percentage, fill=Name.of.company)) + geom\_bar(stat= "identity") + theme\_minimal()

plot(g1)

g1 + ggtitle("COMPANY NAME Vs MEASURE X") + theme(axis.text.x=element\_text(angle=90,hjust=1,vjust=0.5))





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.

ANSWER: IQR range = [5 12]

50% of the data lies between this IQR range

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

ANSWER: It is positively skewed

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?

ANSWER: the new box plot will have a no outlier and the 2.5 data will be in lower whisker region



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

ANSWER: Mode of the dataset lies between 4 and 8

1. Comment on the skewness of the dataset.

ANSWER: Positively skewed with long tail on right 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.

ANSWER:

Median of the dataset is 7

Mode lies between 4 -8

Right skewed so mean > median

50% of the data lies between 5 to 12

Has an outlier with value 25

It is positively skewed

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.)

ANSWER:

x<-1/200 #P(E) = 1/200

y<-1-x # for not occuring the event

#IN 5 ATTEMPTS GETTING ATLEAST 1 MISDIRECTED

p<-1-(y ^ 5) # p = 0.02475 atleast one time occuring the event in 5 independent attempts

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?

ANSWER: Max of P(x) = 0.3 for x=2000

1. Is the venture likely to be successful? Explain

ANSWER: P(x>0) = 0.6 i.e; greater than for P(x<=0) =0.4

HENCE THE BUSINESS WILL BE SUCCESSFUL

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

ANSWER: EXPECTED AVERAGE EARNING = MEAN OF P(x) =0.1667

prob<-c(0.1, 0.1, 0.2, 0.2, 0.3, 0.1)

mean(prob) #mean of P(x) = 0.1667

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

ANSWER: P(x=-2000)+P(x=-1000) = 0.2