**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% |

R CODE: (Copied Data to Excel and saved as Assignment.xls

Data1 = Assignment1 #Saved dataset in excel and imported boxplot(Data1$`Measure X`) #Plotting boxplot to view outliers IQR = IQR(Data1$`Measure X`)

#Calculating IQR

Q = as.numeric(quantile(Data1$`Measure X`))

#For Calculating Q1,and Q3 Q1 = Q[2] Q3 = Q[4]

l = length(Data1$`Measure X`) v = Data1$`Measure X` i=1

while(i< l)

{

if(v[i] > (Q3+1.5\*IQR))

print(v[i]) #0.9136 i=i+1 }

mean = mean(Data1$`Measure X`)

std = sd(Data1$`Measure X`)

var = var(Data1$`Measure X`)

cat("Mean =",mean,"Standard Deviation=",std,"Variance=",var) #Printing Multiple Outputs

Mean = 0.331617

Standard Deviation = 0.189162

Variance = 0.035782

Outlier = 0.9136



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

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

ANS: IQR = 12-5 = 7, this represents the range which contains 50% of the data points.

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

ANS: Right skewed

* 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: 2.5 will be not considered an outlier. The boxplot will start from 0 and send at 20 in

representation.



Answer the following three questions based on the histogram above.

* Where would the mode of this dataset lie?

**ANSWER:** Mode of this dataset lie between 4 and 8.

* Comment on the skewness of the dataset.

**ANSWER:** The dataset is right skewed.

* 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 in boxplot and Mode in histogram.

Histogram provides the frequency distribution so we can see how many times each data point is occurring; however, boxplot provides the quantile distribution i.e., 50% data lies between 5 and 12. Boxplot provides whisker length to identify outliers, no information from histogram. We can only guess looking at the gap that 25 may be an outlier.

* 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.)

**Solution:**

**Given:**

**one in 200 long-distance telephone calls is misdirected.**

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

One in 200 long-distance telephone calls is misdirected

Probability of call misdirecting p = 1/200

Probability of call not Misdirecting = 1 – 1/200 = 199/200

Number of Calls = 5

P(x) = ⁿCₓpˣqⁿ⁻ˣ

n = 5

p = 1/200

q = 199/200

at least one in five attempted telephone calls reaches the wrong number

= 1 – none of the call reaches the wrong number

= 1 – P (0)

= 1 – ⁵C₀ (1/200) ⁰ (199/200) ⁵⁻⁰

= 1– (199/200) ⁵

**= 0.02475**

* 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 |

* What is the most likely monetary outcome of the business venture?

**ANSWER:** P = 0.3 for P (2000). So most likely the outcome is 2000.

* Is the venture likely to be successful? Explain

**ANSWER:** P(x>0) = 0.6, this implies that there is a 60% chance that the venture would yield profits or greater than expected returns. P (Incurring losses) is only 0.2. So, the venture is likely to be successful.

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

**ANSWER:** Weighted average = x\*P(x) = 800. By this the average expected earnings over a long period of time would be 800 (including all the losses and gains over the period of time).

* What is the good measure of the risk involved in a venture of this kind? Compute this m **ANSWER:** P(loss) = P (x= -2000) + P(x=-1000) = 0.2