**Topics: Descriptive Statistics and Probability**

1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
|  | **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% |

X = 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 = 15

Q2 = 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)/15 = 499.07/15

Q2 = 33.27

Q1 = 0.25\*n = 0.25\*15 = 3.75 = 4th observation = 25.41

Q3 = 0.75\*n = 0.75\*15 = 11.25 = 12th observation = 35.00

IQR = 35.00-25.41 = 9.59

Lower extreme = Q1-1.5IQR = 25.41-14.385 = 11.025

Upper extreme = Q3+1.5IQR = 35.00+14.385 = 49.385

Therefore, outlier is 91.36

After removing outlier:

= mean = sum of all numbers / total number of elements

= (24.23+25.53+25.41+24.14+29.62+28.25+25.81+24.39+40.26+32.95+25.99+39.42

+26.71+35)/14 = 407.71/14 = 29.12

=variance = mean of squared deviation score =

= [(24.23-29.12)2+(25.53-29.12)2+(25.41-29.12) 2+(24.14-29.12) 2+(29.62-29.12) 2+(28.25

-29.12) 2+(25.81-29.12) 2+(24.39-29.12) 2+(40.26-29.12) 2+(32.95-29.12) 2+(25.99-29.12) 2+

(39.42-29.12) 2+(26.71-29.12) 2+(35.00-29.12) 2]/14

= (-4.89-3.59-3.71-4.98+0.5-0.87-3.31-4.73+11.14+3.83-3.13+10.30-2.41+5.88)/14

= (23.9121+12.8881+13.7641+24.8004+0.25+0.7569+10.9561+22.3729+124.0996+14.6689+9.7969+

106.09+5.8081+34.5744)/14

= 404.7385 / 14

= 28.91

= standard deviation = square root of variance = = = = 5.38



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.

Ans: Inter-quartile range = Q3 – Q1 = 12 to 5

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

Ans: It’s negative cause majority of observations are on right side to the median

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?

Ans: Here, 25 point is the outlier in the original box-plot, if the value 25 is actually 2.5 then it is in between lower quartile and lower extreme i.e. whisker so it won’t be outlier anymore. Therefore, in the new box-plot there will be no outlier.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Ans: approximately between 4 to 8

1. Comment on the skewness of the dataset.

Ans: The majority of observation are left to the median, therefore skewness of dataset is positive.

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.

Ans: histogram and boxplot both use to explore and present the data in easy and understandable manner. Histogram prefers to determine the underlying probability distribution of data and on the other hand box-plot is more useful while comparing between several datasets

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: one in 200 long-distance telephone calls is misdirected

Probability of call misdirecting p=1/200

Probability of call not misdirecting = 199/200

No. of calls = 5

P(x) = nCxpxqn-x

Here, 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 no.

= 1 - P(0)

= 1 – 5C0(1/200)0(199/200)5-0

= 1 - (199/200)5

= 0.02475

Therefore, the probability that at least one in five calls reaches the wrong no. = 0.02475

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: most likely monetary outcome of the business venture is 2000 as it has maximum probability 0.3

1. Is the venture likely to be successful? Explain

Ans: venture is successful if X is positive. Hence, if X is 1000, 2000 or 3000 probability is 0.2+0.3+0.1 = 0.6 as 0.6 > 0.5 Hence venture likely to be success.

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

Ans: long-term average earning of business ventures = E(X)

E(X) = = $800

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

Ans: Risk involved in a venture

Var(X) = E(X2) = -{E(X)}2

= 2800000 - 8002

= 2160000 (quite high)

SD = $1470

As variability is quite high hence risk is high.