**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% |

ANSWER :

|  |  |  |  |
| --- | --- | --- | --- |
| Name of company | Measure X | (x-mu) | (x-mu)^2 |
| Allied Signal | 0.2423 | -0.09041333 | 0.00817457  1 |
| Bankers Trust | 0.2553 | -0.077413333 | 0.00599282  4 |
| General Mills | 0.2541 | -0.078613333 | 0.00618005  6 |
| ITT Industries | 0.2414 | -0.091313333 | 0.00833812  5 |
| J.P.Morgan & Co. | 0.2962 | -0.036513333 | 0.00133322  4 |
| Lehman Brothers | 0.2825 | -0.050213333 | 0.00252137  9 |
| Marriott | 0.2581 | -0.074513333 | 0.00556715 |
| MCI | 0.2439 | -0.088813333 | 0.00788780  8 |
| Merrill Lynch | 0.4026 | 0.069886667 | 0.00488414  6 |
| Microsoft | 0.3295 | -0.003213333 | 1.03255E-05 |
| Morgan Stanley | 0.9136 | 0.580886667 |  |
| Sun Microsystems | 0.2599 | -0.072813333 |  |
| Travelers | 0.3942 | 0.061486667 |  |
| US Airways | 0.2671 | -0.65613333 | 0.00430511 |
| Warner-Lambert | 0.35 | 00.017286667 | 0.0009882  9 |

# Mean= sum of total values(x)/ total number of values

# = 0.3762

# Standard deviation = sqrt of variance

# = 0.2245

# Variance = (standard deviation ) 2

# nrn = 0.05040



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.
2. What can we say about the skewness of this dataset?
3. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

ANSWERS :

1. Inter-quartile range (IQR) = 12-5 = 7

This implies that the middle 50% of the data lies between 5 and 12.

1. The dataset is right skewed, data is more concentrate towards the left and the right tail is longer.
2. The median will move slightly towards the left

The Q3 will move towards the left

The boxplot will no longer have any outliers.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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.

ANSWERS :

1. The mode of the data set lie will be 4 to 8
2. The data is right skewed, which means the data is more concentrated towards the left side and the the right tale is longer.
3. Median in boxplot and mode will be in histogram

Histogram provides the ‘frequency’ of distribution so that we can find the occurance of data points and however boxplot provides ‘quantile distribution;

Boxplot provides whisker length to identify Outliers, no information from histogram.

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.)

The probability of call being misdirected = 1/200 = 0.005

The probability of call not being misdirected = 1- 0.005 = 0.995

The probability at least one in five attempted calls reaches the wrong number = 1-

The probability that no call reaches the wrong number

= 1- ((0.995)^5)

= 0.024

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?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

# ANSWER :

|  |  |  |
| --- | --- | --- |
| X | P(x) | x.P(x) |
| -2,000 | 0.1 | -200 |
| -1,000 | 0.1 | -100 |
| 0 | 0.2 | 0 |
| 1000 | 0.2 | 200 |
| 2000 | 0.3 | 600 |
| 3000 | 0.1 | 300 |

# P = 800

# 1 . the most likely monetary outcome of the business venture is :x = 2000 with highest probability is 0.3.

# 2 . the venture=re is likely to be successful because the probability of the venture having a return on investment is 0.6(0.2+0.3+0.1).

# 3 . the long-term average earning of business venture is (0.1)(-2000)+(0.1)(1000)+(0.2)(0)+(0.3)(2000)+(0.1)(3000)=800

# 4 . Standard deviation is [1] 0.08164966