**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 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?

**Ans:** (i) **IQR** = 12.5-5

= 7.5

IQR indicates the how the data is spread between the first and third quartiles.

(ii) The given dataset is positively skewed, since more values are on the right tail.

(iii) If the data point was actually 2.5, then outlier in the data set will be replaced by the

new data point and the new data set will not contain outlier. Also, the median value of the data set will change, move to left side. That is, the median will become lesser compared to the previous case.



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.

**Ans:** (i) The data set is bimodal with modes 5 and 7.

(ii) The data set is positively skewed.

(iii)Both plot gives information about the data set. The boxplot explains descriptive measures

like median, range, inter-quartile range, quartiles, skewness etc, where histogram explains

mode, sample size, skewness etc. We can also fit theoretical distribution to the given data

set using histogram. Both plots give idea about outliers.

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:**  Probability (calls misdirecting) = 1/200 = p

Probability (calls not misdirecting) = 1-(1/200)

= 199/200 = q

n = 5

Let X be the event that calls being misdirected.

Probability (at least one in five attempted telephone calls reaches the wrong number) = P (X = 1)

P (X = 1) = 5C1p1q4

= 0.0245

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

**Ans:** (i) $2000 is the most-likely outcome with probability 0.3.

(ii) The probability of success of this venture is 0.2 + 0.3 + 0.1 = 0.6

The venture is likely to be successful, since the probability of success is greater than 0.5.

(iii) Average = Σ xi\*P(xi)

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

= 800

The long-term average earning of this venture is $800.

(iv) The best measure of risk involved in this venture is variance.

Variance = E(X2) - (E(X))2

E(X2) = Σ(Xi2 \* P(Xi))

= 2800000

Var(X) = 2800000 – 8002

= 2160000

Standard Deviation = 1469.694

The risk is high since the value of variance is high.