**Topics: Descriptive Statistics and Probability**

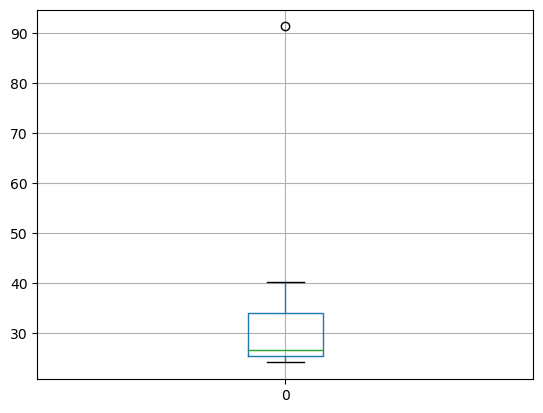
1. Look at the data given below. Plot the data, find the outliers and find out

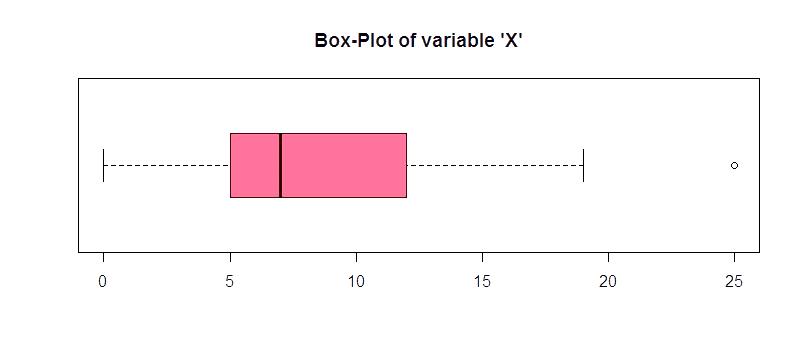
|  |  |
| --- | --- |
| **Name of company** | **Measure X** |
| Allied Signal | 24.23% |
| Bankers Trust | 25.53% |
| General Mills | 25.41% |
| ITT Industries | 24.14% |
| J.P.Morgan & Co. | 29.62% |
| Lehman Brothers | 28.25% |
| Marriott | 25.81% |
| MCI | 24.39% |
| Merrill Lynch | 40.26% |
| Microsoft | 32.95% |
| Morgan Stanley | 91.36% |
| Sun Microsystems | 25.99% |
| Travelers | 39.42% |
| US Airways | 26.71% |
| Warner-Lambert | 35.00% |

Answer:

* Creating the list as ‘ls’ in Jupyter
* Converting the list into data frame as ‘df’
* Plotting the box plot as df. Boxplot ()
* The outlier is 91.36
* Calculating the mean as df. mean ()
* Calculating the standard deviation as df. Describe ()
* Mean = 33.271333
* Std (standard deviation) σ = 16.945401
* = = 287.1466151
* variance is √std = 4.1164

​



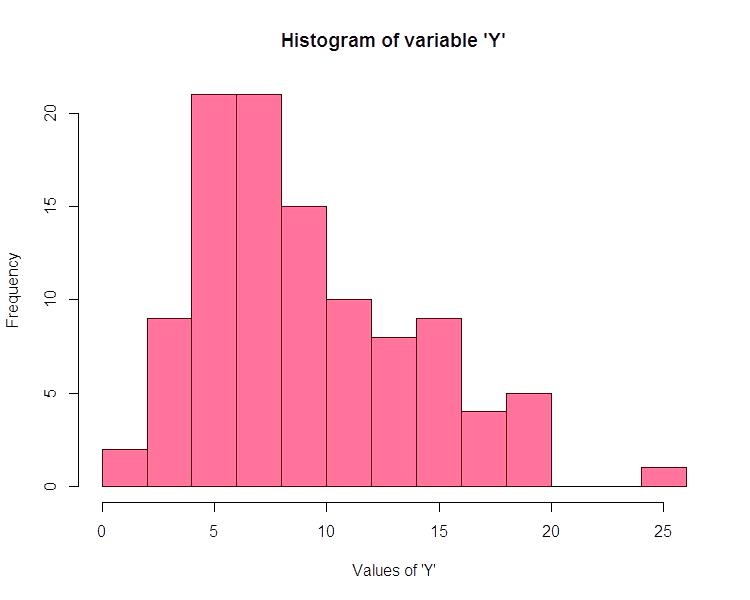


Answer the following three questions based on the box-plot above.

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?

Answer:

* Quartile 1 Upper End (Q1) = 5
* Quartile 3 Lower End (Q3) = 12
* Interquartile range (IQR) = (Q3 – Q1) = 12 – 5 = 7
* IQR implies that 50% of data set values lies in the region of IQR i.e. around 7
* The data is skewed towards the outlier (25), it is negatively skewed as the mean is influenced by the presence of outlier
* If the point 25 would be 2.5 then 2.5 will fall in the range of box plot. There would be no outlier present in the data so the data set will be symmetrical
* If the point 25 would be 2.5 then mean = median = mode



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.

Answer:

* The mode of the data set lies between the range [ 4, 8] (Approx)
* The data set is positively skewed as the mean of this data set is 8.85 ≈ 9, mode is 6.5 and median is 13 so, median < mean < mode.
* Both have the outliers as 25
* Both the histogram and boxplot have high frequency around 6 as value
* In histogram the mean has very less influence of outlier due to high frequency of low values so it is positively skewed while in box plot the outlier has high influence towards mean so it is negatively 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:

This problem is a binomial distribution, where each attempted telephone call is a Bernoulli trial with a probability of success (getting misdirected)

p and a probability of failure (not getting misdirected) q, where q=1−p.

In this case:

p (probability of success, i.e., a call getting misdirected) is  because one in 200 calls is misdirected.

q (probability of failure, i.e., a call not getting misdirected) is 1− .

Now, the probability of at least one success (at least one call being misdirected) in five attempts can be calculated using the complement rule (i.e., 1 minus the probability of zero successes).

The probability mass function of the binomial distribution is given by:

P (X = k) = × × (1 − )

where:

n is the number of trials (5 attempts),

k is the number of successes (at least one call being misdirected),

p is the probability of success on a single trial,

q is the probability of failure on a single trial.

The probability of at least one success in five attempts is given by:

P (X ≥ 1) = 1 − P (X=0)

Now, let's calculate it:

P (X ≥ 1) = 1 − 0 × × (1 -

P (X ≥ 1) = 1−

P (X ≥ 1) ≈ 0.0249

So, the probability that at least one in five attempted telephone calls reaches the wrong number is approximately 0.0249 or 2.49%.

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:

i) Most Likely Monetary Outcome:

The most likely monetary outcome is the value of x with the highest probability. In this case, the value 2000 has the highest probability (0.3), so the most likely monetary outcome is $2000.

(ii) Likelihood of Success:

To determine if the venture is likely to be successful, we need to consider the probabilities associated with positive returns. Positive returns occur when x is greater than 0.

P (Success)=P (X>0) = P (1000) + P (2000) + P (3000)

P(Success)=0.2+0.3+0.1=0.6

The venture has a 60% chance of being successful based on this probability distribution.

(iii) Long-Term Average Earnings:

The long-term average earnings (μ) can be calculated as the weighted sum of the possible outcomes:

= × P

μ = (−2000⋅0.1) + (−1000⋅0.1) + (0⋅0.2) + (1000⋅0.2) + (2000⋅0.3) + (3000⋅0.1)

μ=−200−100+0+200+600+300=800

The long-term average earnings are $800.

(iv) Measure of Risk:

A common measure of risk is the standard deviation (σ), which provides information about the variability of returns.

= √ × P

σ= ( × 0.1) + (× 0.1) + (× 0.2) + (× 0.2) + ( × 0.3) + ( × 0.1)

= [(-2,800)^2 \* 0.1] + [(-1,800)^2 \* 0.1] + [(-800)^2 \* 0.2] + [200^2 \* 0.2] + [1,200^2 \* 0.3] + [2,200^2 \* 0.1]

= [7,840,000 \* 0.1] + [3,240,000 \* 0.1] + [640,000 \* 0.2] + [40,000 \* 0.2] + [1,440,000 \* 0.3] + [4,840,000 \* 0.1]

= 784,000 + 324,000 + 128,000 + 8,000 + 432,000 + 484,000

σ ≈ √2160000 ≈1470

The standard deviation is $1600, providing a measure of the risk involved in the venture.

In summary:

(i) The most likely monetary outcome is $2000.

(ii) The venture is likely to be successful with a 60% chance of positive returns.

(iii) The long-term average earnings are $800.

(iv) The measure of risk, represented by the standard deviation, is $1600.