**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: Solved using Jupyter Note Book (Python)- File attached

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive statistics | Mean (µ) | Standard Deviation (σ) | Variance (σ2) |
| Jupyter Code | x**.**mean() | x**.**std() | x**.**var() |
| Answer | 33.27133333333333 | 16.945400921222028 | 287.1466123809524 |



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.

Answer:

Upper Quartile (UQ): 12

Lower Quartile(LQ): 5

Inter Quartile Range (IQR)= UQ-LQ

= 12-5

= 7

The Value 7 Is the Median

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

Answer:

Right-Skewed/Positively Skewed

Median is towards the left side

Data is not normally distributed

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?

Answer:

No outlier will be observed.

No change in the median Value

Changes in IQR

Data will be normally distributed



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

Answer: The mode of this data lies in between 4 to 8 approximately.

1. Comment on the skewness of the dataset.

Answer:Right/Positively Skewed (Mean>Median>Mode)

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.

Answer:

* Both data are right-skewed
* Both data have outliers.
* Box Plot: The median visualized
* Histogram: mode is more visible.

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:

One in 200 long-distance telephone calls are getting misdirected.  
probability of call misdirecting = 1/200

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

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

Number of Calls = 5 n = 5 p = 1/200 q = 199/200

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

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

P(x) = (nCx) (p^x) (q^n-x)

# nCr = n! / r! \* (n - r)!

P(1) = (5C1) (1/200)^1 (199/200)^5-1

P(1) = 0.0245037

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?

Answer:

2000$, since the probability is maximum in comparison to others i.e., 0.3.

1. Is the venture likely to be successful? Explain

Answer:

Yes

p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8

80% chances for this venture to be making a profit

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

Answer:

 Expected value = Sum (X \* P(X)) = 800$ which means on an average the returns will be + 800$

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

Answer:

This kind depends on the Variability in the distribution.

Higher Variance means more chances of risk

Var (X) = E(X^2) –(E(X))^2 = 2800000 – 800^2 = 2160000