**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% |

**Ans:**

**x=pd.Series([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])**

**#box plot**

**sns.boxplot(x)**

**#mean , standard deviation ,variance**

**x.mean()**

**x.std()**

**x.var()**

Mean: **33.27133333333333**

Standard deviation: **16.945400921222028**

Variance: **287.1466123809524**

The following is the outlier in the boxplot. i.e. **Morgan Stanley 91.36%**



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(IQR)= Q3-Q1.

The above diagram is said to be upper whisker length.

The formula for upper whisker lengths is = **Q3+1.5(IQR)**

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

**Ans:** The above diagram is said to be **Right-skew.**

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:  In that case there would be no Outliers on the given dataset because of the outlier the data had positive skewness it will reduce, and the data will normal distributed



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans:** The mode of the data set is lies **“Right skewed ”.**  median towards the left side it is not normal distribution.

1. Comment on the skewness of the dataset

**Ans**:. The skewness of the data set is lies on “**Right-skewed** ” **.**  which is 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.

**Ans:**  they both are Right-skewed and both have outliers the median can be easily visualized in box plot where as 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.)

**Ans:**  n=5

p=1/200

Bi=binom(n=5,p=1/200)

#p(x<1)

p=1-Bi.cdf(0)

p

* **0.02475124687812502**

The probability that at least one in five attempted telephone calls reaches the wrong number is “**0.02475124687812502”**

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**: The most likely monetary outcome of the business venture is **2000** and the probability is **0.3**

1. Is the venture likely to be successful? Explain

**Ans:**  Yes, when we consider 0-3000 the venture is likely successful and the probability is

**0.2+0.2+0.3+0.1=0.8**. there are 80% chances for this venture making profit

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

**Ans:** The long-term average is Excepted value =Sum(X\*p(x))= 800$ which means on an average the return will be **800$**

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

**Ans:**  The good measure of the risk involved in a venture of this kind depends on the variability in the distribution. Higher variance mean more chances of risk var(x) =E(X^2)-(E(x)^2)

* **2800000-800^2=2160000**