**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% |

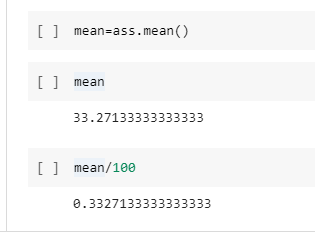
From python we get the values of µ,σ,σ2 as follows:

µ= 0.3327

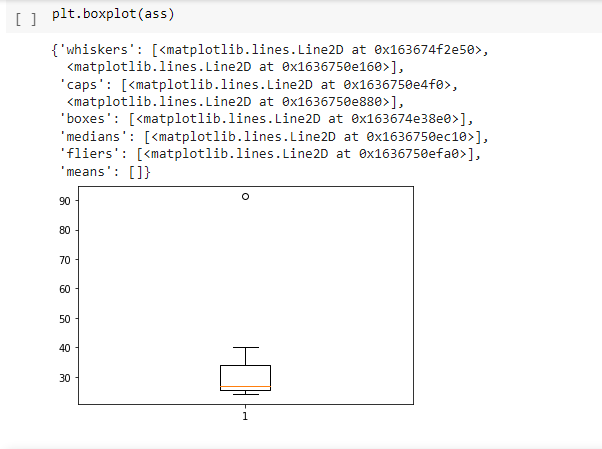
σ= 0.1694

σ2 = 0.0287

Hence, from box plot is it clear that there is outlier on higher side which is 91.36%







2.



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:-** The inter-quartile range (IQR) of datasets is 5/12. The value implies that 50% data is in between 5-12.

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

**Ans:-** The dataset has outlier on higher side, so the skewness of data set is positively skewed.

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:-** The new boxplot does not have outlier because value 2.5 is in lower whisper. It will reduce the right skewness of the data.

3.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**Ans:-** The mode of this data approx. set lies between 5 to 10 and highly Between 4-8

1. Comment on the skewness of the dataset.

**Ans:-** Skewness of dataset is positively skewed.

From this data we can say 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:-** If histogram & boxplot are plotted then we can easily get the outliers, because in boxplot it is shown & in histogram we can get it from its frequency.

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:** IF 1 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 |

**Ans:-** given: probability distribution of return on a certain business venture , to the nearest $1000

Solution : • E(X) = ∑X . P(X)

• E(X²) = ∑X² . P(X)

• Var (X) = E(X²) - { E(X) }²

• SD = √Var

|  |  |  |  |
| --- | --- | --- | --- |
| X | P(X) | E(X)=X.P(X) | E(X²) = X² . P(X) |
| -2000 | 0.1 | -200 | 400000 |
| -1000 | 0.1 | -100 | 100000 |
| 0 | 0.2 | 0 | 0 |
| 1000 | 0.2 | 200 | 200000 |
| 2000 | 0.3 | 600 | 1200000 |
| 3000 | 0.1 | 300 | 900000 |
|  | 800 |  | 2800000 |

1. What is the most likely monetary outcome of the business venture?

**Ans:-** $2000

1. Is the venture likely to be successful? Explain

**Ans:-** Venture is successful if X is +ve hence if X is 1000, 2000 or 3000 probability is

0.2+0.3+0.1 = 0.6 as 0.6>0.5 hence venture likely to be successful

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

**Ans:-** Average earning = expected values **= ∑ Xi \* P(Xi) = $ 800**

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

**Ans:-** Risk involved in a venture var (X) = E(X²)  - { E(X) }² =   2800000 -   800²

=2160000 (quite height) SD= √var ≈$ 1470 as variability is quite high hence risk is high.