**Topics: Descriptive Statistics and Probability**

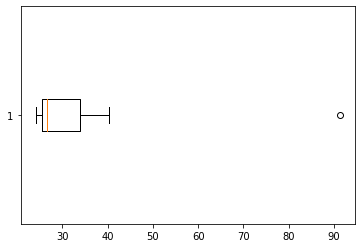
1. Look at the data given below. Plot the data, find the outliers and find out

|  |  |
| --- | --- |
| **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% |

import matplotlib.pyplot as plt

plt.boxplot(x, vert = False)

plt.show()



Outlier is “Morgan Stanley” with 91.36%

import pandas as pd

x = [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, 35.00]

df = pd.DataFrame(x)

df.mean()

df.std()

df.var()

**Output :**

1. Mean ( = 33.271333 %
2. Standard deviation ( = 16.945401 %
3. Variance ( = 287.146612 %



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.

A) First quartile range (Q1) = 5

Second quartile Range (Q2) = 7

Third quartile Range (Q3) = 12

Inter Quartile Range (IQR) = Q3 – Q1 = 12 – 5 = 7

IQR implies that the 50% of lies between Q1 and Q3

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

A) The data 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?
2. In that case the data has no Outliers and the box-plot needs to be reconstructed as there is a slight reduction in the skewness of that data.



Answer the following three questions based on the histogram above.

(i) Where would the mode of this dataset lie?

A) The mode of the above dataset lies approximately between 4 to 8

(ii) Comment on the skewness of the dataset.

1. The dataset the Positively Skewed

(iii) 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.

A) There is an outlier of the value 25 and both are Positively 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.)
2. Total number of calls = 200

P(Misdirected call) = 1/200

P(Correct Calls) = 1-1/200 = 199/200

P(misdirected calls in 5 attempts) = P(C)5 = (199/200)5

We want to find the probability that at least one call is misdirected in five attempts.

P(at least one misdirected call in 5 attempts) = 1 - P(no misdirected calls in 5 attempts)

P(at least one misdirected call in 5 attempts) = 1 - (199/200)5 = 0.0248

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 |

(i) What is the most likely monetary outcome of the business venture?

A) The most likely outcome is the return that has the highest probability. In this case, the most likely outcome is **$2,000**, which has a probability of 0.3.

(ii) Is the venture likely to be successful? Explain

1. If success is defined solely by making a profit, then

P(Success) = P(x ≥ 1000) = P(x = 1000) + P(x = 2000) + P(x = 3000)

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

(iii) What is the long-term average earning of business ventures of this kind? Explain

1. The long-term average earning of business ventures of this kind is equal to its **expected value (EMV)**

**EMV = sum(x\*p(x))**

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

**EMV = $800**

Therefore, over the long term, the average earning of business ventures with this probability distribution is expected to be $800.

(iv) What is the good measure of the risk involved in a venture of this kind? Compute this measure

A) The good measure of the risk involved in a venture of 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 – 8002 = 2800000 – 640000 = $2160000