**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% |

**Code :** fig = plt.figure(figsize=(10, 4));

ax = sns.barplot(x=names, y=data, edgecolor='black');

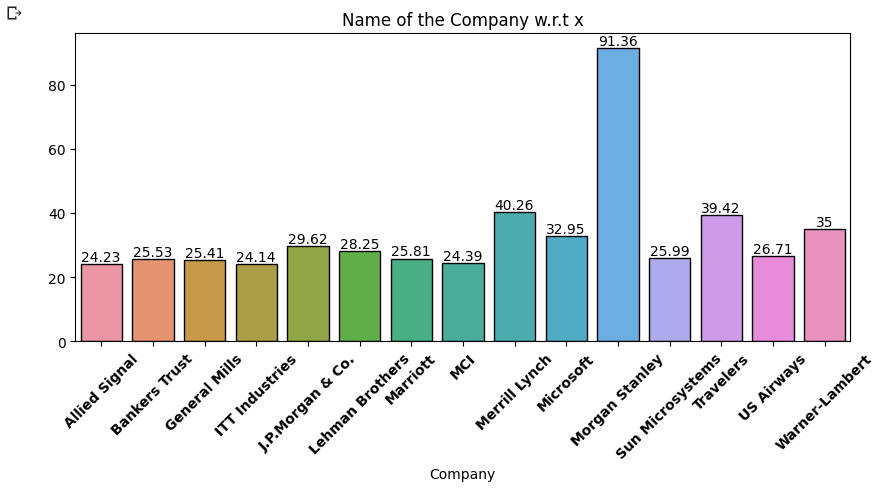
plt.xticks(rotation = 45, fontweight = 'bold');

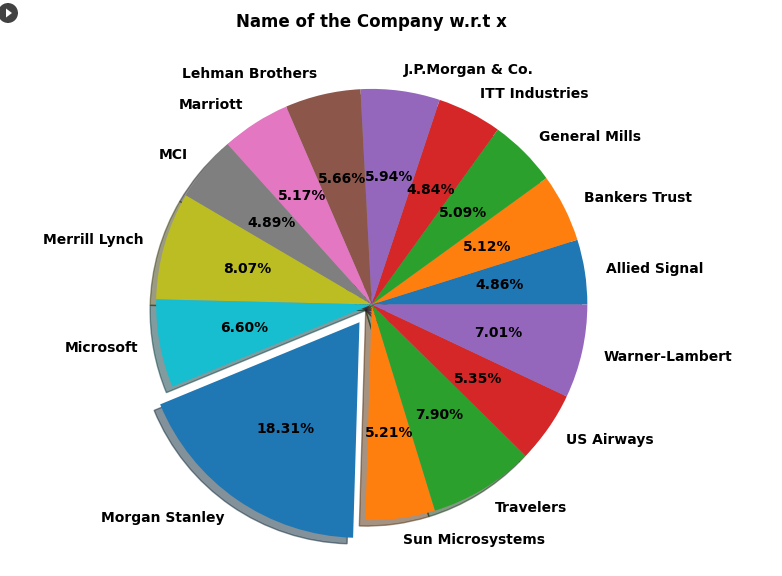
for i in ax.containers:

     ax.bar\_label(i,)

plt.xlabel('Company', color='black');

plt.title('Name of the Company w.r.t x');





**Code:** fig = plt.figure(figsize=(14, 7));

plt.pie(data, labels = names, autopct = '%1.2f%%', shadow = True,

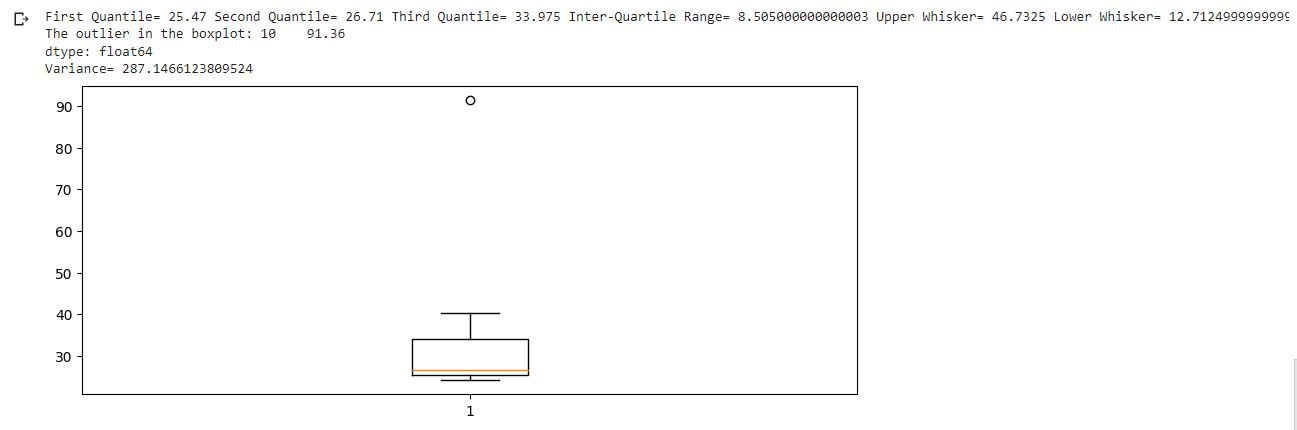
        explode = [0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.0,0.1,0.0,0.0,0.0,0.0],

        textprops = {'size':'medium',

                   'fontweight':'bold',

                   'color':'black'});

plt.title('Name of the Company w.r.t x', fontweight='bold');



**Code:** fig = plt.figure(figsize=(10, 4));

plt.boxplot(data);

LQ = np.quantile(data,0.25)

UQ = np.quantile(data,0.75)

median = np.median(data)

IQR = UQ - LQ;

LE = LQ - 1.5 \* IQR

UE = UQ + 1.5 \* IQR

print('First Quantile=', LQ, 'Second Quantile=', median, 'Third Quantile=', UQ,

      'Inter-Quartile Range=', IQR, 'Upper Whisker=', UE, 'Lower Whisker=', LE)

#Outliers = data[(data <= LE) | (data >= UE)]

#print('The outlier in the boxplot:',Outliers)

Outliers = data[(data >= UE)]

print('The outlier in the boxplot:',Outliers)

print("Variance=",data.var(), 'Mean=', data.mean(), 'Standard deviation=', data.std())

First Quartile= 25.47, Second Quartile = 26.71, Third Quartile = 33.975, Inter-Quartile Range= 8.505000000000003, Upper Whisker= 46.7325, Lower Whisker= 12.712499999999995

The outlier in the boxplot: 10 91.36

Variance= 287.1466123809524, Mean= 33.27133333333333, Standard Deviation= 16.945400921222028



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:** Approximately 🡪 Lower Quartile= 5; Upper Quartile = 12; Median = 7

(Inter-Quartile Range) IQR = Upper Quartile – Lower Quartile= 12 – 5 = 7

Median is also called as second quartile range.

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

**Answer:** The dataset given is positively skewed data or right skewed data. As skeweness is present hence dataset is not having normal distribution.

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:** if 25 is actually 2.5 then dataset would not have any outliers and data would probably 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 dataset would lie in between 5 to 10.

1. Comment on the skewness of the dataset.

**Answer:** data is right skewed or positively skewed, where 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:** Histogram and box plot both have outliers and both plots are positively skewed. We can easily visualize median in box plot and mode in histogram.

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 is misdirected.

We have to find the probability that at least one in five attempted telephone calls reaches the wrong number.

Probability of long distance telephone call is misdirected (p) = 1/200

Probability of long distance telephone call is not misdirected (q) = 1 - 1/200

= 200 – 1 /200

=199/200

Probability of at least one in five attempted telephone calls reaches the wrong number 🡪

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

Where n = 5 (number of calls attempted)

At least 1 in 5 telephone calls reaches the wrong number that is no telephone calls reaches the correct number 🡪 P(x) = 1 – P(0)

P (0) 🡪 0 in 5 telephone calls reaches the correct number.

= 1 - 5C0 p0 q5-0

= 1 - 5C0 (1/200) 0 (199/200) 5-0

= 1 - 1 \* 1 \* 0.9752487531218

= 0.0247512468781

= 0.024

That is the probability that at least one in five attempted telephone calls reaches the wrong number is 2.4% or 2.5%

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:** As the probability is 0.3 for 2000, more than others, thus most likely monetary outcome of the business venture = 2000$.

1. Is the venture likely to be successful? Explain

**Answer:** Ventureis likely to be successful if x is positive. Thus, when x is 1000, 2000 or 3000, probability is 0.2 + 0.3 + 0.1 = 0.6, which is > 0.5, that is more than 50% long term average gives positive numbers.

Hence, venture is likely to be successful.

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

**Answer:** Long-term average means expected value 🡪

Long-term average of business ventures is 🡪 Expected value (E(X)) = Sum (x\*P(x)) ∑

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

= 800$

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:** Risk involved in a venture of this kind depends on variance, means we have to calculate variance.

Variance = ∑(x – μ)2 ⋅ P(x)

(-2000 – 800)2 ⋅ 0.1 = 784000

(-1000 – 800)2 ⋅ 0.1 = 324000

(0 – 800)2 ⋅ 0.2 = 128000

(1000 – 800)2 ⋅ 0.2 = 8000

(2000 – 800)2 ⋅ 0.3 = 432000

(3000 – 800)2 ⋅ 0.1 = 484000

= 784000 + 324000 + 128000 + 8000 + 432000 + 484000

= 2160000

Here variance is quite high, hence risk is also high.