**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 -:**

**Measure\_x=pd.Series([24.23,25.53,25.41,24.14,29.62,28.25,25.81,24.39,40.26,32.95,91.36,25.99,39.42,26.71,35.00])**

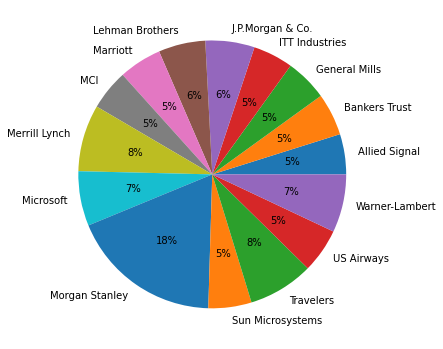
**name=['Allied Signal','Bankers Trust','General Mills','ITT Industries','J.P.Morgan & Co.','Lehman Brothers', 'Marriott','MCI','Merrill Lynch','Microsoft','Morgan Stanley','Sun Microsystems','Travelers','US Airways', 'Warner-Lambert']**

**# Pie Plot**

**plt.figure(figsize=(6,8))**

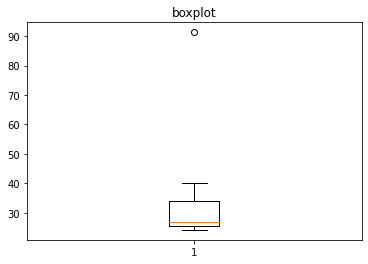
**plt.pie(Measure\_x,labels=name,autopct='%1.0f%%')**

**plt.show()**



**plt.boxplot(Measure\_x)**

**plt.title("boxplot")**



**Measure\_x.mean()**

**33.27133333333333**

**Measure\_x.median()**

**26.71**

**Measure\_x.var()**

**287.1466123809524**

**Measure\_x.std()**

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

**Second Quartile Range is the Median Value**

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

**Answer -:** **Right-Skewed median is towards the left side it is not 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 -:**  **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?

**Answer -:** **The mode of this data set lie in between 5 to 10 and**

**approximately between 4 to 8 .**

1. Comment on the skewness of the dataset.

**Answer-:** **Right-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** -: **They both are right-skewed and both have outliers**

**the median can** **be easily visualized in box plot where**

**as in 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 -:**

**probability of call misdirecting   = 1/200**

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

**Number of Calls = 5**

**P(x) = ⁿCₓpˣqⁿ⁻ˣ**

**n = 5**

**p = 1/200**

**q = 199/200**

**at least one in five attempted telephone calls reaches the wrong number**

**= 1  -  none of the call reaches the wrong number**

**= 1  - P(0)**

**= 1   -  ⁵C₀(1/200)⁰(199/200)⁵⁻⁰**

**= 1  -  (199/200)⁵**

**= 0.02475**

**probability that at least one in five attempted telephone calls reaches the wrong number = 0.02475**

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 |

**Answer -:**

|  |  |  |  |
| --- | --- | --- | --- |
| **X** | **P(X)** | **E(X)= X . P(X)** | **E(X²) = X² . P(X)** |
| **-2,000** | **0.1** | **-200** | **400000** |
| **-1,000** | **0.1** | **-100** | **100000** |
| **0** | **0.2** | **0** | **0** |
| **1000** | **0.2** | **200** | **200000** |
| **2000** | **0.3** | **600** | **12000000** |
| **3000** | **0.1** | **300** | **900000** |
| **Total** |  | **800** | **28000000** |

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

**Answer -:** **Most likely monerate outcome of** **the business venture is $2000 as it has maximum Probability 0.3**

1. Is the venture likely to be successful? Explain

**Answer -: Yes, the probability that the venture will make more than 0 or a profit**

**p(x>0)+p(x>1000)+p(x>2000)+p(x=3000) = 0.2+0.2+0.3+0.1 = 0.8 this states that there is a good 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 -:** **The long-term average is 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 -:** **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 – 800^2**

**= 2160000**