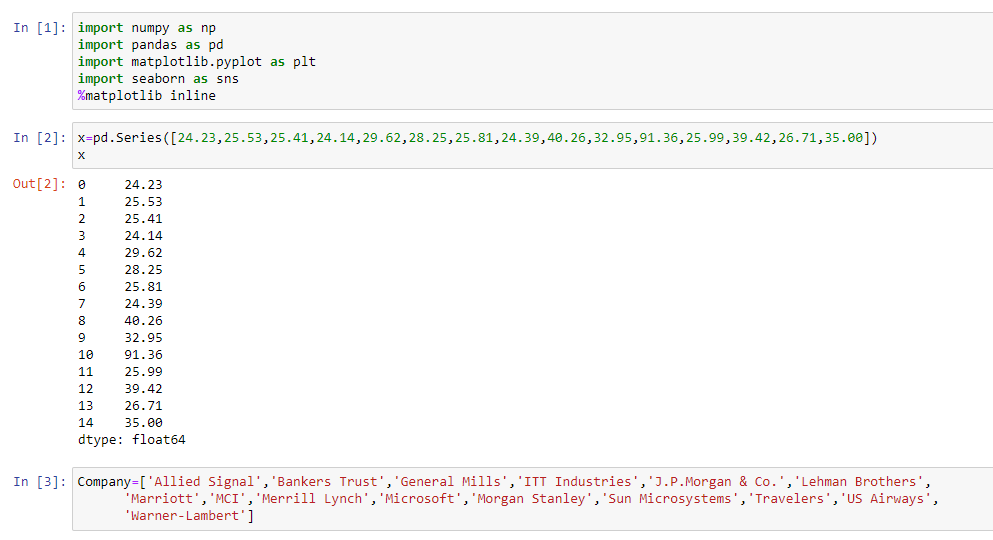
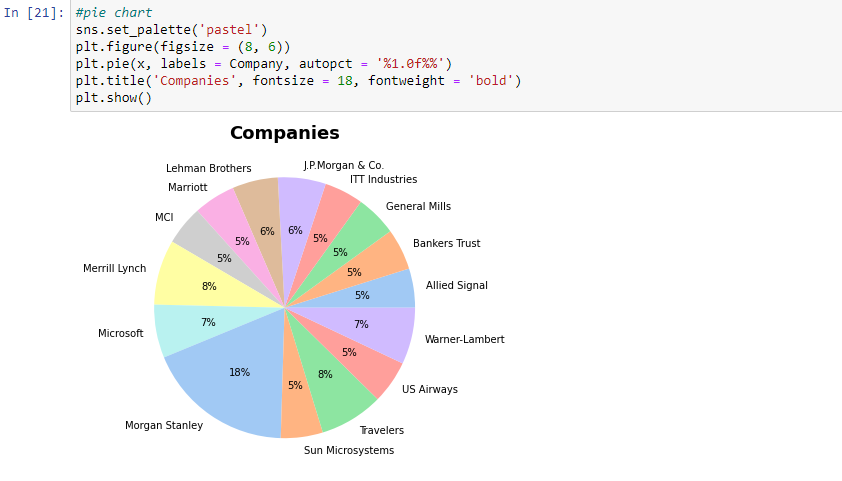
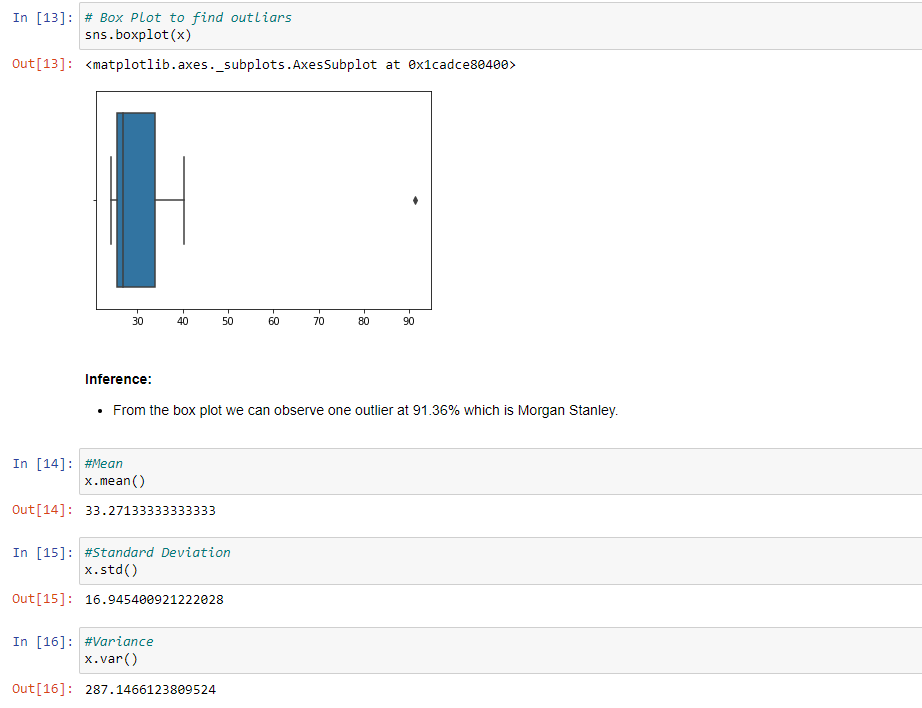
**Topics: Descriptive Statistics and Probability**

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 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.

Interquartile range (IQR) is the amount of spread in the middle 50% of a dataset. In other words, IQR describes the difference between the 3rd Quartile and 1st Quartile (75th percentile and 25th percentile).

IQR = Q3 - Q1 = 12 – 5 = **7**

Here the IQR value implies that **Mean > median**

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

The tail of the dataset is skewed towards right and peak is towards left, so it is a **right side skewed** or positively skewed. (i.e.: The data constitute higher frequency of high value)

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?

To see if box-plot is affecting with the new data point 2.5, let’s check the boundaries for the outlier. Outlier will fall outside of this range,

[Q1-1.5(IQR), Q3 +1.5(IQR)] =5\*10.5, 12+10.5 = **5.5, 22**

**5.5, 22-** these are the boundaries for the outlier. So, any datapoint less then 5.5 and greater than 22 will be an outlier. Hence, **datapoint 2.5 will be an outlier.**



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

The mode lies on the 7 on the X – axis

1. Comment on the skewness of the dataset.

The data is Right side skewed

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.

***Inferences from the histogram and boxplot***

* From the plots we can say, it is **Right side skewed or positively skewed.**
* There is an **outlier** at the data point 25
* The median falls at data point 7

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.)

**Solution:**

one in 200 long-distance telephone calls is misdirected

probability of call misdirecting, p = 1/200

      Probability of call not Misdirecting, q= 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 |

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

The most likely monetary outcome of the business venture: x= 2,000

with the highest probability of 0.3

1. Is the venture likely to be successful? Explain

The venture is likely to be successful, because (x= 1,000) + (x= 2,000) + (x= 3,000) = 0.2 + 0.3 + 0.1 = 0.6

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

|  |  |  |
| --- | --- | --- |
| x | P(x) | x ⋅ P(x) |
| -2,000 | 0.1 | (-2000).(0.1) = -200 |
| -1,000 | 0.1 | (-1000).(0.1) = -100 |
| 0 | 0.2 | (0).(0.2) = 0 |
| 1000 | 0.2 | (1000).(0.2) = 200 |
| 2000 | 0.3 | (2000).(0.3) = 600 |
| 3000 | 0.1 | (3000).(0.1) = 300 |

Add the last column *x***⋅***P*(*x*) to find the long term average or expected value: (0.1)(−2,000) + (0.1)(−1,000) + (0.2)(0) + (0.2)(1,000) + (0.3)(1,000) + (0,1)(3,000)= 800.

We say *μ* = 800.

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

* The good measure of the risk means std dev and var.
* To calculate the standard deviation (*σ*) of a probability distribution, find each deviation from its expected value, square it, multiply it by its probability, add the products, and take the square root.
* To find the standard deviation, add the entries in the column labeled

(*x* – *μ*)2*P*(*x*) and take the square root.

|  |  |  |  |
| --- | --- | --- | --- |
| x | P(x) | x ⋅ P(x) =  *μ* | (*x* – *μ*)2*P*(*x*) |
| -2,000 | 0.1 | (-2000).(0.1) = -200 | (-2000 – 200)2 . 0.1 = 484000 |
| -1,000 | 0.1 | (-1000).(0.1) = -100 | (-1000 – 100)2 . 0.1= 121000 |
| 0 | 0.2 | (0).(0.2) = 0 | (0 – 0)2 . 0.2 = 0 |
| 1000 | 0.2 | (1000).(0.2) = 200 | (1000 – 200)2 . 0.2 = 128000 |
| 2000 | 0.3 | (2000).(0.3) = 600 | (2000 – 600)2 . 0.3 = 588000 |
| 3000 | 0.1 | (3000).(0.1) = 300 | (3000 – 300)2 . 0.1 = 729000 |

Variance, σ2  = sum of all ((x – μ)2 ⋅ P(x))

=484000+ 121000+ 128000 + 588000 + 729000 = **2050000**

Standard deviation, *σ* = √2050000 =**1431. 78**