**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% |

* Mean=33.27
* Variance=287.14
* Standard deviation=16.94
* **def** find\_outliers(data):

*# Calculate the first and third quartiles*

q1 **=** np**.**percentile(data, 25)

q3 **=** np**.**percentile(data, 75)

*# Calculate the interquartile range (IQR)*

iqr **=** q3 **-** q1

*# Find the lower and upper bounds*

lower\_bound **=** q1 **-** 1.5 **\*** iqr

upper\_bound **=** q3 **+** 1.5 **\*** iqr

*# Identify the outliers*

outliers **=** [x **for** x **in** data **if** x **<** lower\_bound **or** x **>** upper\_bound]

**return** outliers

*# Test the function with a sample data set*

* data **=** [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]
* outliers **=** find\_outliers(data)

print(outliers)

* [91.36]



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=> IQR = Q3-Q1 = 12-5 = 7 (Approximately).

It means 50% of data points lie in the range of 5 and 12.

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

Answer=> The dataset is positively skewed.

Tail is found extending towards right side of the curve.

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 outliers the data had positive skewness it will reduce and the data will normal distributed.

3.



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=> Positive 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 positive skewedand bothe 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 p = 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 - (199/200)⁵

= 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=>

Returns on a certain business venture, to the nearest $1,000, are known to follow the following probability distribution

1. `X ` ``P(X)` ` E(X)= X . P(X)` `E(X²) = X² . P(X)`
2. `-2000 ` `0.1` `-200` `400000`
3. `-1000` `0.1` `-100 ` `100000`
4. `0 ` `0.2` `0` `0`
5. `1000` `0.2` `200` `200000 `
6. `2000` `0.3 ` `600` `1200000`
7. `3000` `0.1` `300` `900000`
8. `800` `2800000`

E(X) = ∑X . P(X)

E(X²) = ∑X² . P(X)

Var (X) = E(X²) - { E(X) }²

SD = √Var

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

Answer=> most likely monetary outcome of the business venture is $ 2000 as it has maximum Probability 0.3.

1. Is the venture likely to be successful? Explain

Answer=> P(x>0) = 0.6, implies there is a 60% chance that the venture would yield profits or greater than expected returns. P(Incurring losses) is only 0.2.

So the venture is likely to be successful.

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

Answer => Weighted average = x\*P(x) = 800. This means the average expected earnings over a long period of time would be 800(including all losses and gains over the period of time.

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.

P(loss) = P(x= -2000)+P(x=-1000)=0.2. So the risk associated with this venture is 20%