[Company name]  [Company address]

SET+1+descriptive+stast+probability

**Topics: Descriptive Statistics and Probability**

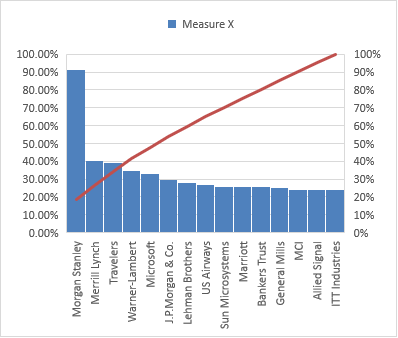
1. Look at the data given below. Plot the data, find the outliers and find out

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

**ANS:-**

**A)Data representation:-**

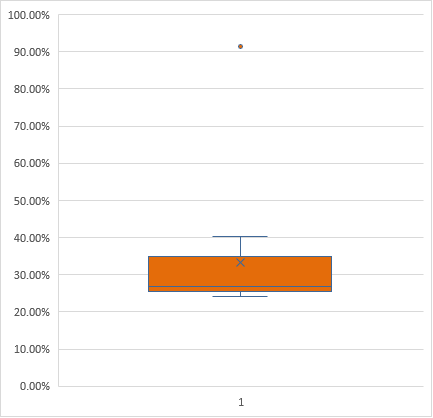
**PARETO CHART**



**PIE CHART**

**B)Outlier Detection:-**

**Outlier is define by using box plot**



From Excel by using QUARTILE.INC

Table No1

|  |  |
| --- | --- |
| Minimum | 0.2414 |
| Q1 | 0.2547 |
| Median | 0.2671 |
| Q3 | 0.33975 |
| Maximum | 0.9136 |

Table No2

|  |  |
| --- | --- |
| Mean | 33.27% |
| Range | 0.6722 |
|  |  |
| IQR | 0.08505 |
|  | 0.127575 |
| Lower cut off | 0.127125 |
| Higher cut off | 0.467325 |

From above observation and blox **plot lower cut off and higher cut off for normal data is 12% and 46% respectively** so above this and below this is consider as an **outlier**. In our data there is one value which is above **higher cut off so it is consider as an outlier i.e 91.36**

**A)MEAN():-** From excel

= **33.27%**

**B)Variance():-**

**σ2 = ∑ f (m − x̅)2 / n**

**= 0.0268**

**C)Standard deviation():-**

**= Sq.root of variance**

**=Sq.root of(0.0268)**

**=0.163708**



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.

**ANS:- IQR=Q3-Q1**

**Q3=**75% of data

**Q1=**25% of data

**From above box plot, Q3=13, Q1=5**

Therefore, **IQR=Q3-Q1**

**=13-5**

**IQR=8**

**IQR = Q3 – Q1 This tells us how spread out the middle 50% of values are in a given dataset.**

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

**ANS:- The data is positively skeweed as there is one outlier at higher side.**

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?

**Ans:-** Then data becomes **left skeweed** as outlier is shifted below lower limit.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

**ANS:-Between 5 to 6**

1. Comment on the skewness of the dataset.

**ANS:- The data is positively skwened.**

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.

**ANS:- Then box plot will clearly shows value of outliers,where histogram shows distribution of data.**

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.)

ANS:- IF 1 in 200 long-distance telephone calls are getting misdirected.

probability of call misdirecting = 1/200

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

The probability for at least one in five attempted telephone calls reaches the wrong number

Number of Calls = 5

n = 5

p = 1/200

q = 199/200

P(x) = at least one in five attempted telephone calls reaches the wrong number

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ

P(x) = (nCx) (p^x) (q^n-x) # nCr = n! / r! \* (n - r)!

P(1) = (5C1) (1/200)^1 (199/200)^5-1

P(1) = 0.0245037

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 |

E(X) =Sum X.\*P(X) | E(X^2) =X^2\*P(X)

-200 | 400000

-100 | 100000

0 | 0

200 | 200000

600 | 1200000

300 | 900000

Total: 800 | 2800000

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

Ans: The most likely monetary outcome of the business venture is 2000$

As for 2000$ the probability is 0.3 which is maximum as compared to others

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

Ans: 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

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

Ans: The long-term average is Expected value = Sum (X \* P(X)) = 800$ which means on an average the returns will be + 800$

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

Ans: 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