**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) PFA ipynb for reference…





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.
2. What can we say about the skewness of this dataset?
3. If it was found that the data point with the value 25 is actually 2.5, how would the new box-plot be affected?

(i)ans: Inter-quartile range =(q3-q1)=(12-5) = 7.

(ii)ans: Right skewed data

(iii)ans: if value 25 is 2.5 then distribution have no outliers, mean value are also change and data distribution slightly moves to left side and becomes normally distributed.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?
2. Comment on the skewness of the dataset.
3. 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.

i)ans: mode would be this dataset is 4,5,6,7,8.

Ii)ans: Right skewed

iii)ans: Both distributions are right skewed and have outliers, in box plot we can easily identify mean but histogram easily identify mode.

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)

Probability of misdirected calls(p) = 1/200

Probability of not misdirected calls(q) = 1-(1/200) =199/200

P(x) = ⁿCₓ pˣ qⁿ⁻ˣ  [x=1,n=5]

So, substituting n, x, p, q values in above formula will get probability of at least one in five attempted calls.

P(1) = 5C1\*(1/200)\*(199/200)^(5-1) = 0.0245.

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?
2. Is the venture likely to be successful? Explain
3. What is the long-term average earning of business ventures of this kind? Explain
4. What is the good measure of the risk involved in a venture of this kind? Compute this measure

i)ans : most likely monetary outcome is = 2000. 2000 probability is 0.3 which is greater than others.

ii) ans:

venture will be successful if profit is positive = p(x=1000)+p(x=2000)+p(x=3000)

= 0.2+0.3+0.1 = 0.6

Which means 60% venture likely to be successful.

iii)ans: long-term average earning of business ventures = summation(x\*p(x))

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

=800$.

iv)ans: var(x) = E(x^2)-{E(x)}^2

E(x) is summation of x\*p(x) = 800

E(x^2)= summation(x^2\*p(x)) = 400000+100000+200000+1200000+900000

= 2800000

Var(x) = 2800000 –(800)^2 =2160000

Std =sqrt(2160000) =1470.

Variance is high so, risk is high.