**Topics: Descriptive Statistics and Probability**

* 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:-) Ans is in jupyter notebook file name basic statics level-2



Answer the following three questions based on the box-plot above.

* What is inter-quartile range of this dataset? (please approximate the numbers) In one line, explain what this value implies.

ANS: IQR is the range between upper quartile (Q3) and lower quartile (Q1)

IQR= Q3-Q1= 12-5 = 7

50% of the data lies between IQR

* What can we say about the skewness of this dataset?

ANS:-) this dataset has positive skewness.

* 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: There will be no outlier if the value of 25 was actually 2.5. Subsequently,

and Whether the median shifts or not depends on the size of the data.

It will reduce the right skewness of the data.



Answer the following three questions based on the histogram above.

* Where would the mode of this dataset lie?

ANS:-) The mode will be lie between 4 and 10 because majority of the entry in

this range but this is just an assumption. The 2 bars of the same height

doesn’t indicate mode every time.

* Comment on the skewness of the dataset.

ANS:-) dataset have positive skewness

* 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: There is an outlier of the value 25 and both the plot has positive

skewness

* 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: one in 200 long-distance telephone calls is misdirected

=> 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 - ⁵C₀(1/200)⁰(199/200)⁵⁻⁰

= 1 - (199/200)⁵

= 0.02475

So the probability that at least one in five attempted

telephone calls reaches the wrong number =

0.02475

* 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 |

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

ANS:-) $2000 as it has the highest probability of occurrence

* Is the venture likely to be successful? Explain

ANS:-) if Success == positive returns as a measure

Then there is a 60% probability that the venture would be successful

(0.3+0.2+0.1=0.6=>0.6\*100=>60%)

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

ANS:-) (-2000\*0.1)+(-1000\*0.1)+(0\*0.2)+(1000\*0.2)+(2000 \*0.3)+(3000\*0.1)=800

the long-term average earning for these type of ventures would be around $800

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

ANS:-) A good measure to evaluate the risk would be variance and standard deviation

of the variable x

Var = 3500000

Sd = 1870.83

The large value of standard deviation of $1870 is considered along with the

avarage

returns of $800 indicates that this venture is highly risky

Formula used:

df[“x”].std()

df[“x”].var()