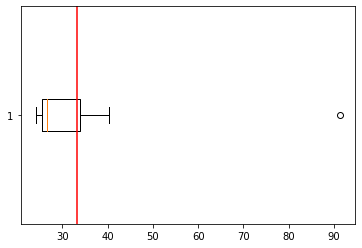
**Topics: Descriptive Statistics and Probability**

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% |

= 33.2713, = 16.9453, = 287.1466, Outlier is Morgan Stanley = 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.

* IQR = 12 – 5 = 7. IQR contains 50% of data used to plot boxplot.

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

* From boxplot we can see data below median is concentrated on **left** of boxplot. So, we can say that data is **‘skewed right’**.

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?

* If value 25 is actually 2.5 then the boxplot **will not** have any outlier and will have less slight change in **skewness value**, might even be considered as **normally distributed**. ‘**Median**’ will also be slightly shifted to right.



Answer the following three questions based on the histogram above.

1. Where would the mode of this dataset lie?

* Mode of dataset lies at values **5** and **7**.

1. Comment on the skewness of the dataset.

* From histogram, we can see that data is mostly concentrated on left-side of plot as well as we have an outlier on right-side of plot. So, we can say that dataset is **skewed right** .

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.

* Boxplot shows us **median** as well as **distribution** of data is situated in given range. Whereas histogram gives us the frequency with which data occurs and why we have **median** toward **left** despite having such wide range.

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.)

* Probability that call is misdirected = **1/200**

Probability that call is not misdirected = 1-(1/200) = **199/200**

For probability that at least one in five attempted telephone calls reaches the wrong number, we use binomial distribution formula.

**P(x) = nCx .px(1 − p) n−x, n = 5 and x = 1**

P(x) = 5C1.P5(1 − p)5−1 = (5! / 4!1!) \* (1/200)1 \* (199/200)4

P(x) = (5 \* 1994) / 2005

**P(x) = 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?

* **2000**

1. Is the venture likely to be successful? Explain

* **Yes**, as probability of making profit is ‘0.2+0.2+0.3+0.1 = 0.8’ which is higher than probability of having loss that is ‘0.1+0.1 = 0.2’.

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

* Long-term average earning of business ventures is equal to **expected** **value** of probability function. Expected value is given as,

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

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

Risk involved in a venture depends on variance in data. Variance in data is given by,

Var(x) = E(x2) – [E(x)]2 = ∑x2.P(x) – [∑x. P(x)]2

**E(x) = 800**

E(x2) = (4000000\*0.1) + (1000000\*0.1) + 0 + (1000000\*0.2) + (4000000\*0.3) + (9000000\*0.1) = 400000 + 100000 + 0 + 200000 + 1200000 + 900000

**E(x2) = 2800000**

**Var(x) = 2800000 – 640000 = 2160000**

**Std = √(Var) = √2160000 = 1469.69**

As **variance and standard deviatio**n is **high**, we can say that **risk** involved in venture is also **high**.