8- What is Zscore?

Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of standard deviations from the mean.

→ If a z-score is 0, it indicates that the data points are identical to the mean score.

→ A Zscore of 1.0 would indicate a value that is one standard deviation from the mean.

→ Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean.

1

In finance, Z-scores are the measures of an observation's variability and can be used by traders to help determine market volatility. The Z-Score is also sometimes known as the Altman Z-score.

$$z = \frac{x - \mu}{\sigma}$$

Benifits

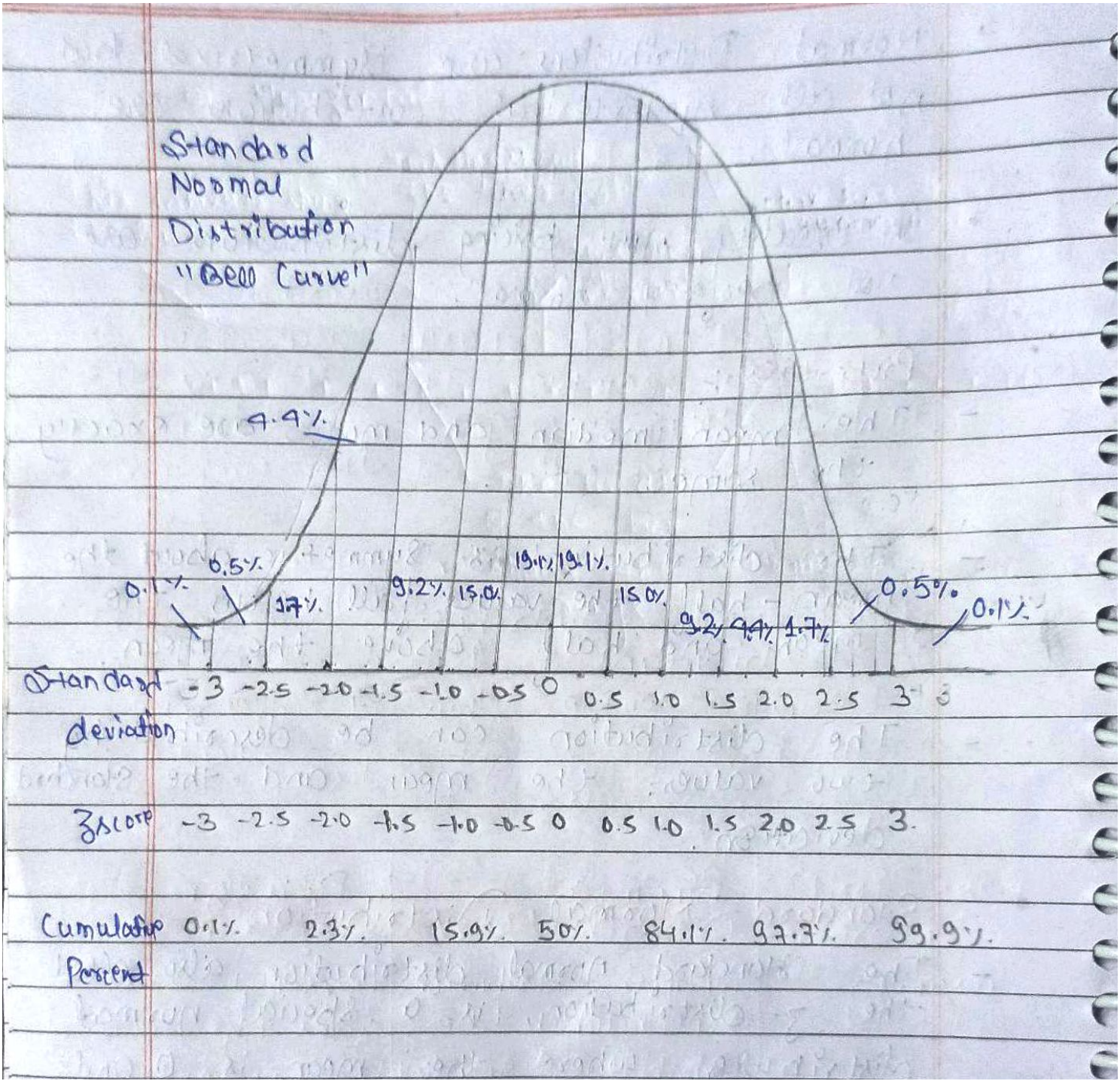
→ Understand where a data point fits into a distribution

→ Compare observation between dissimilar variables

→ Identify Outliers

→ Calculate probabilities and percentiles using the standard normal distribution.

2


Standard
Normal
Distribution
"Bell Curve"
4.4%
0.1%
0.5%
17%
9.2%
15.0%
19.1%
19.1%
15.0%
9.2%
4.4%
1.7%
0.5%
0.1%
Standard deviation -3 -2.5 -2.0 -1.5 -1.0 -0.5 0 0.5 1.0 1.5 2.0 2.5 3
Z score -3 -2.5 -2.0 -1.5 -1.0 -0.5 0 0.5 1.0 1.5 2.0 2.5 3.
Cumulative Percent 0.1% 2.3% 15.9% 50% 84.1% 97.7% 99.9%

Interquartile Range

The quartiles of a ranked set of data values are 3 points which divides the data into exactly four equal parts, each of part a comprising of quarter data.

→ Q_1 → Defined as the middle number between the smallest number and the median of the data set.

→ Q_2 → The median of the data.

→ Q_3 → The middle value between the median and the highest value of the data set.

IQR

→ The Inter Quartile Range IQR tells us the range where the bulk of the value lie.

→ The inter Quartile range is calculated by subtracting the first quartile from the third quartile.

$$IQR = Q_3 - Q_1$$

Use;

1. Unlike range, IQR tells where the majority of data lies and is thus preferred over range.
2. IQR can be used to identify outliers in a data set.
3. Gives the central tendency of the data.

Ex: 1, 19, 2, 6, 5, 9, 12, 27, 18, 2, 15

Total - 15

1, 2, 5, 6, 7, (9) 12, 15, 18, 19, 27

$Q_2 = 9$

$Q_1 = 5$

$Q_3 = 18$

IQR $= Q_3 - Q_1 = 13$

Qs. What are Outliers and Extreme Values?

Outliers

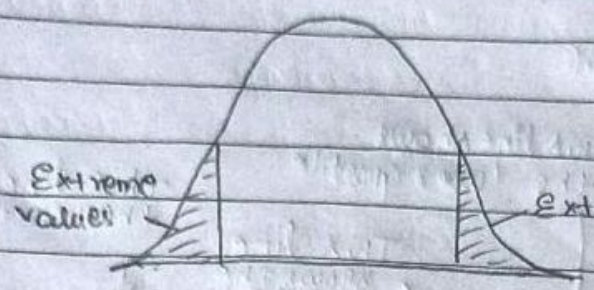
→ In simple terms, an Outlier is an extremely high or extremely low data point relatively to the nearest data point and the rest of the neighboring co-existing value in a data graph or dataset you are working with.

→ Outliers are the extreme values that stand out greatly from the overall pattern of values in a dataset or graph.

Extreme Value

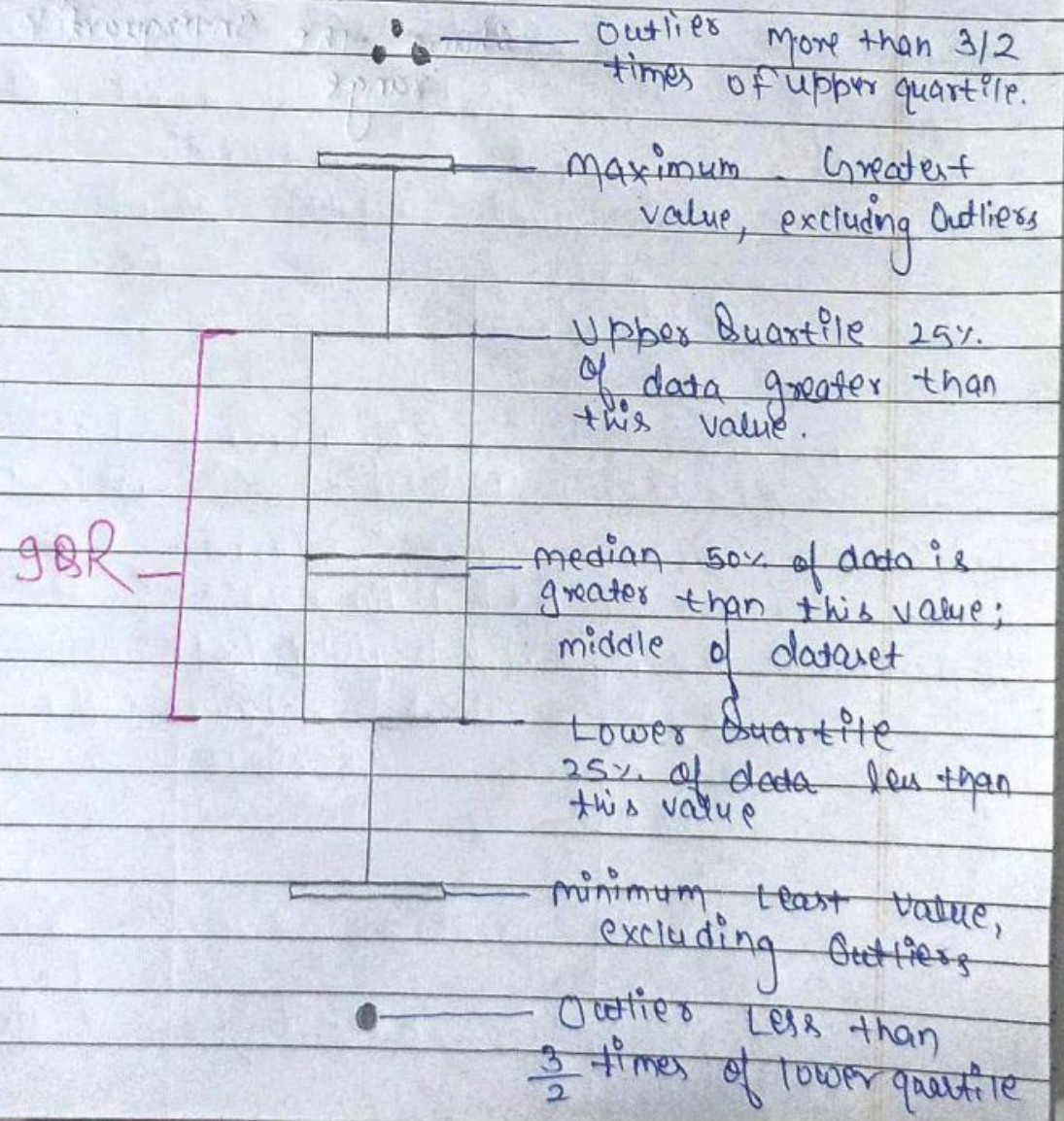
An extreme value is either very small or very large value in a probability distribution.

These extreme values are found in the tails of a probability distribution. (i.e. the distribution tends extremities).

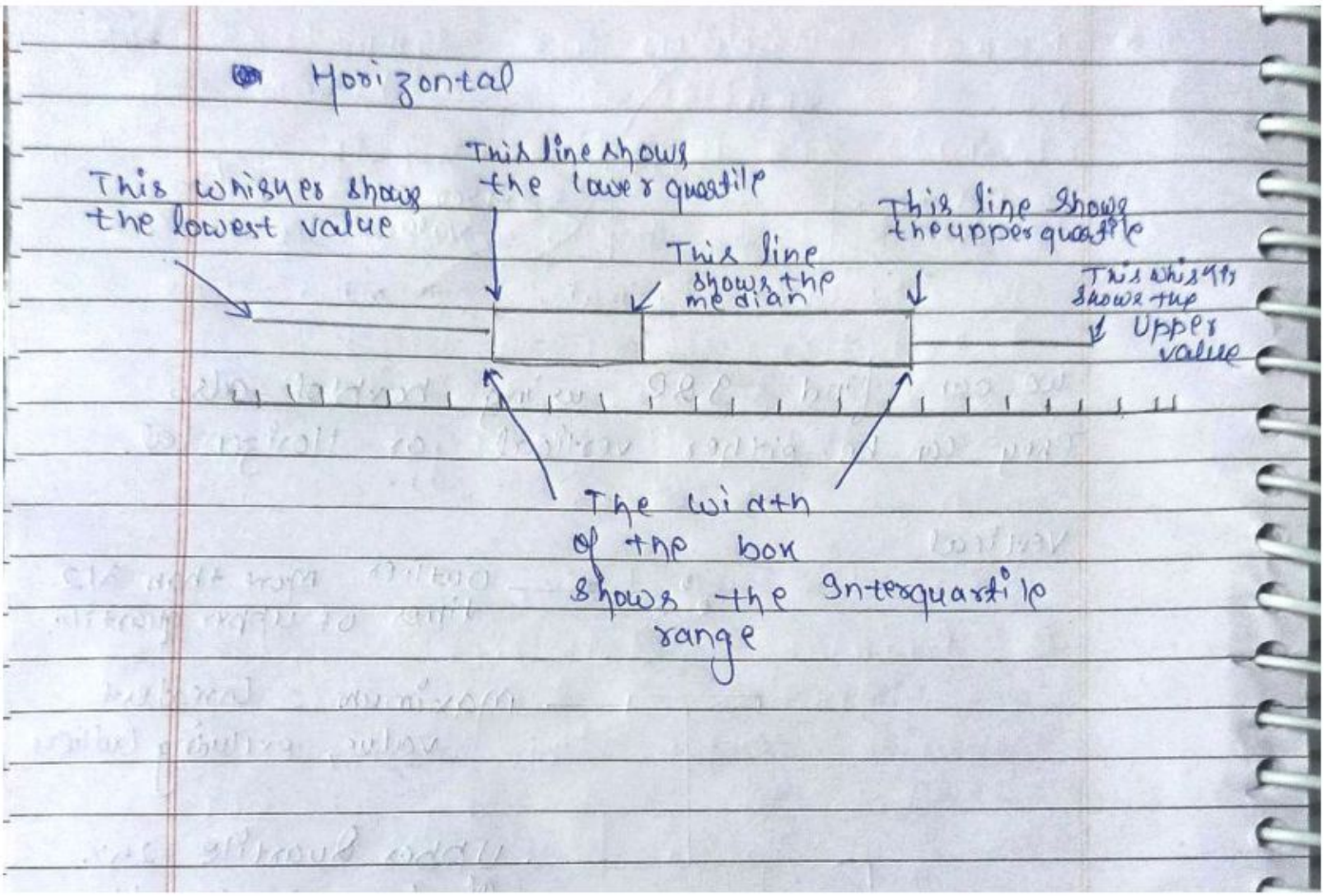


We can find IQR using boxplot also.
They can be either vertical or Horizontal.

Vertical



- Horizontal


This whiskers show the lowest value
This line shows the lower quartile
This line shows the median
This line shows the upper quartile
This whisker shows the upper value
The width of the box shows the interquartile range