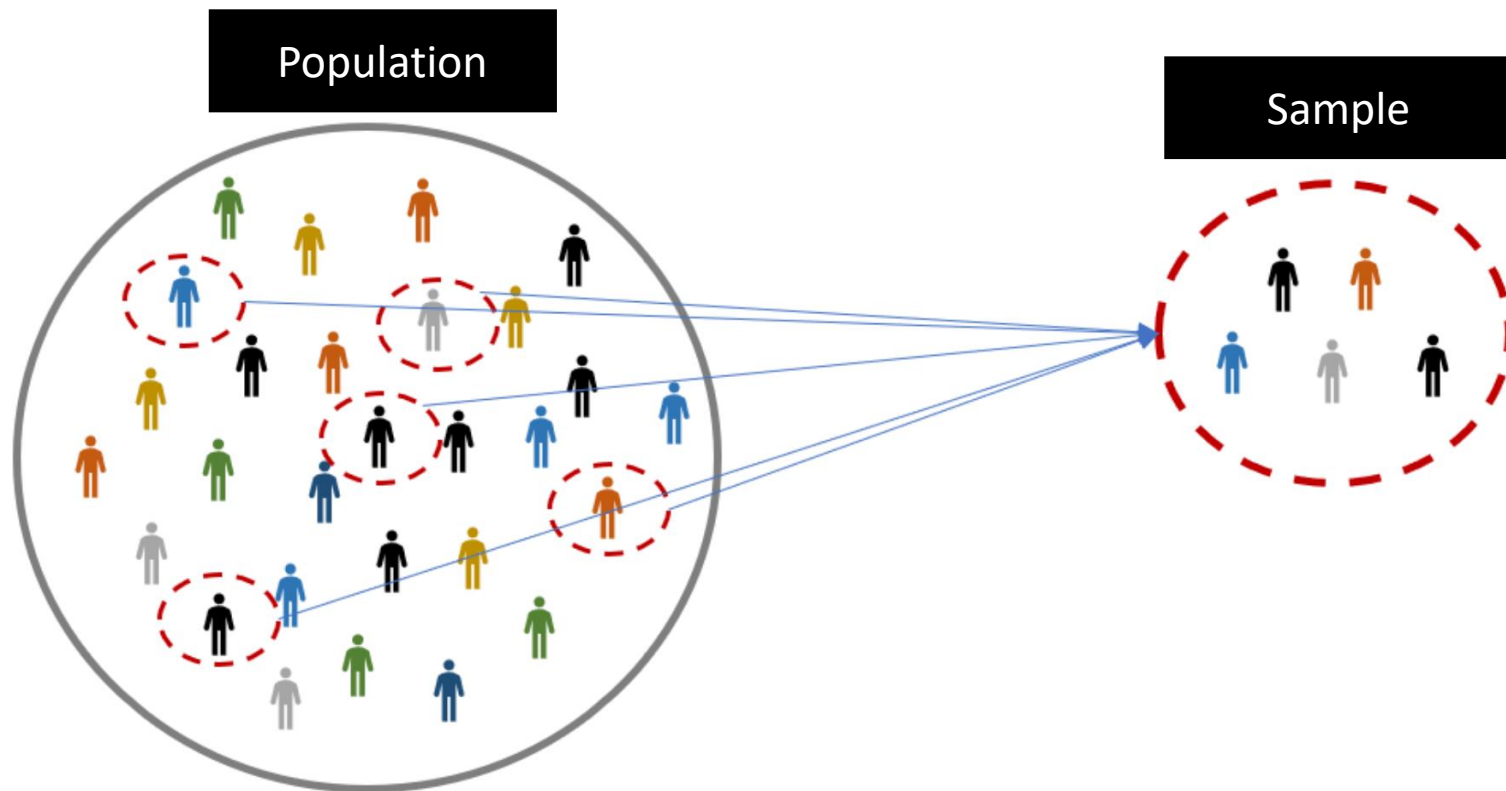


- Population
- Sample
- Standard Deviation
- Histogram
- Skewness
- Normal Distribution / Gaussian Distribution
- Log Normal Distribution
- Z Scores
- Outlier Removal using Python

Population & Sample

A population is the entire group that you want to draw conclusions about. A sample is the specific group that you will collect data from. The size of the sample is always less than the total size of the population.

Population & Sample



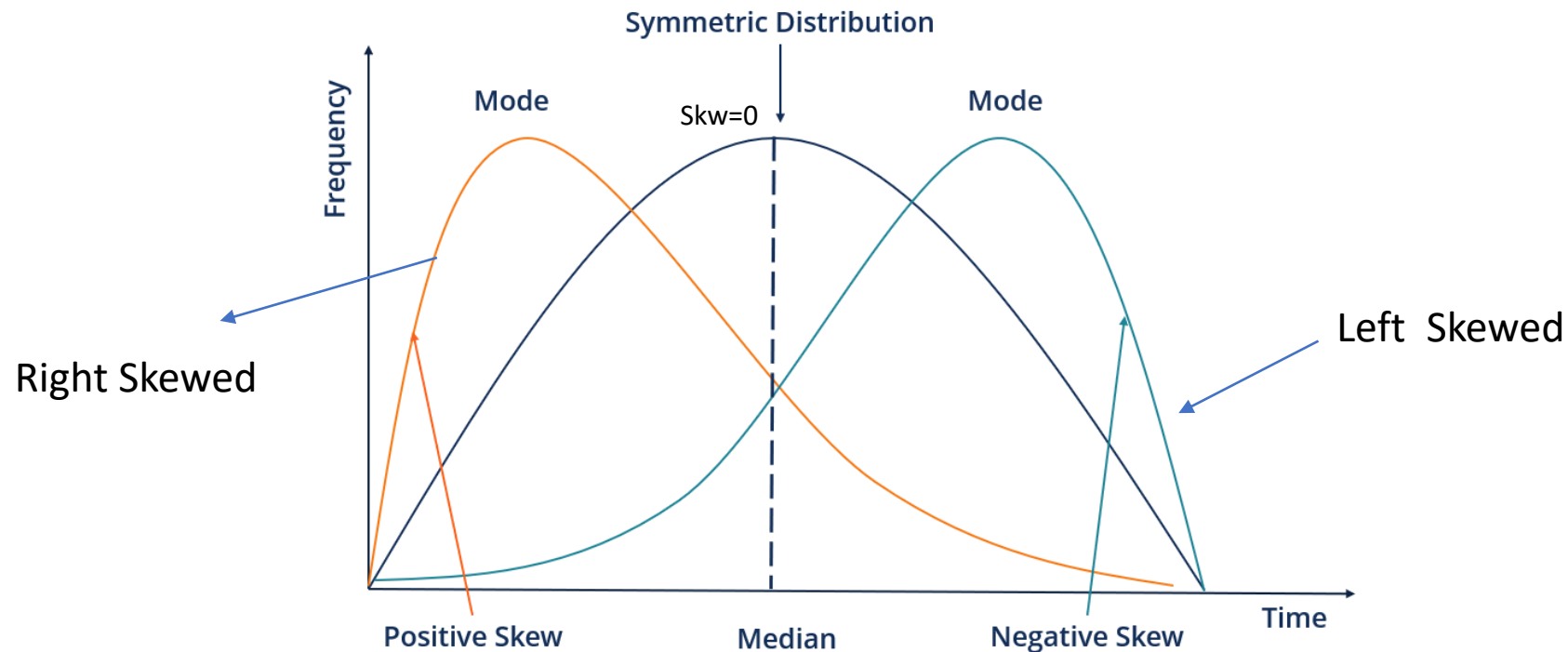
Standard deviation

In statistics, the standard deviation is a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range.

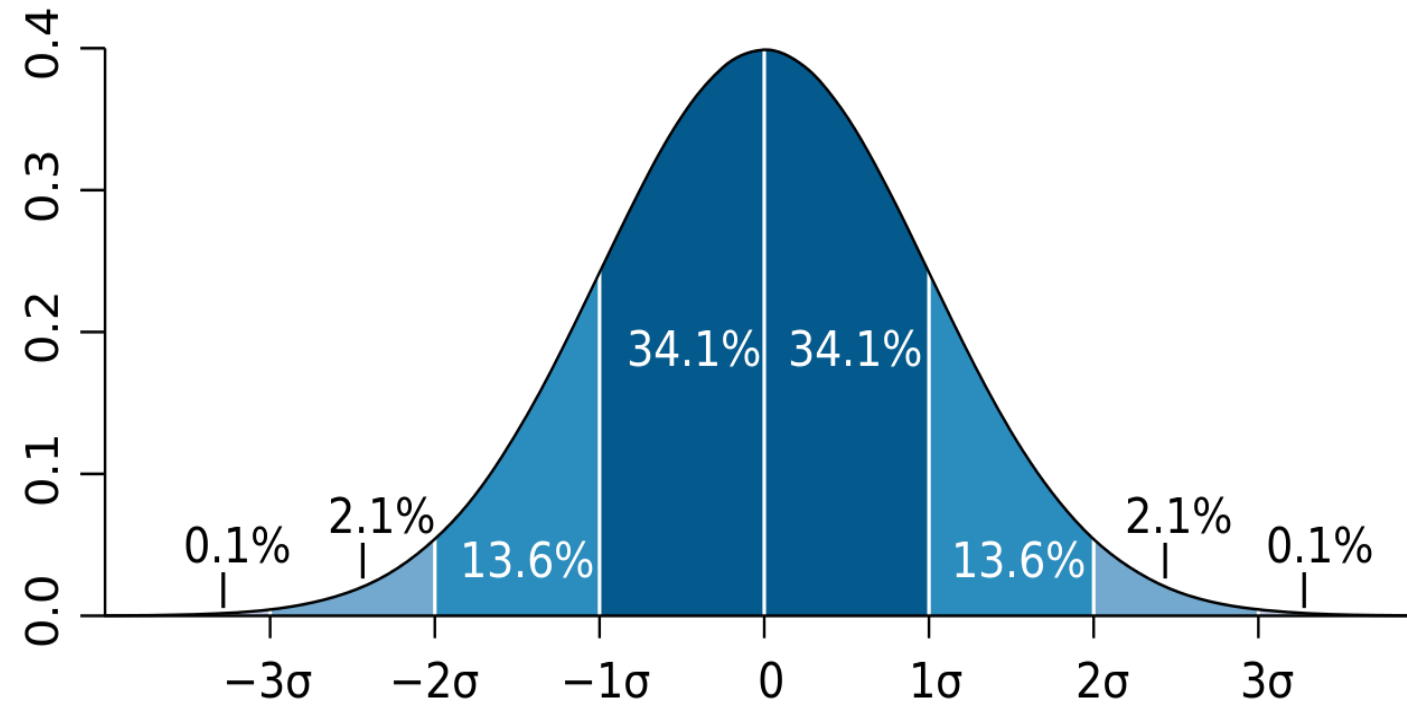
$$SD = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

Skewness

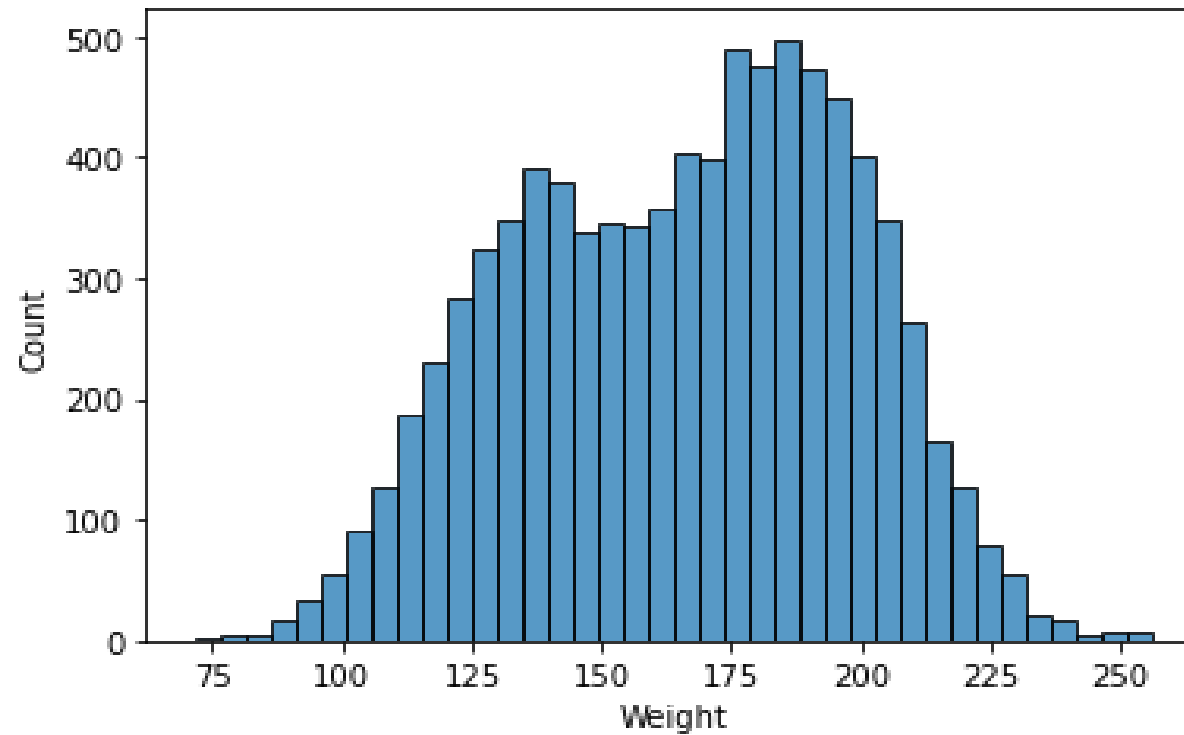
Skewness refers to a distortion that deviates from the symmetrical bell curve, or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed



Standard deviation

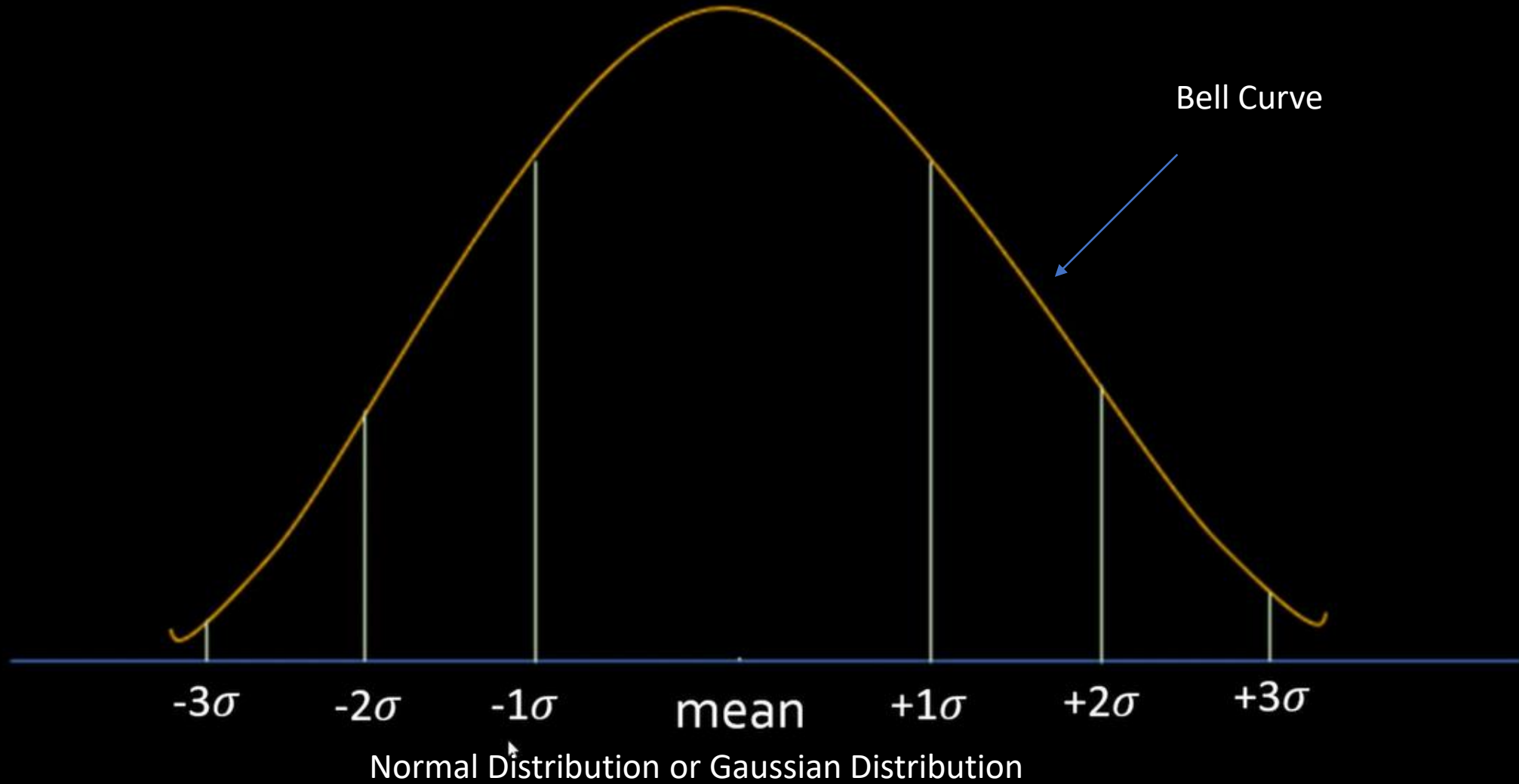


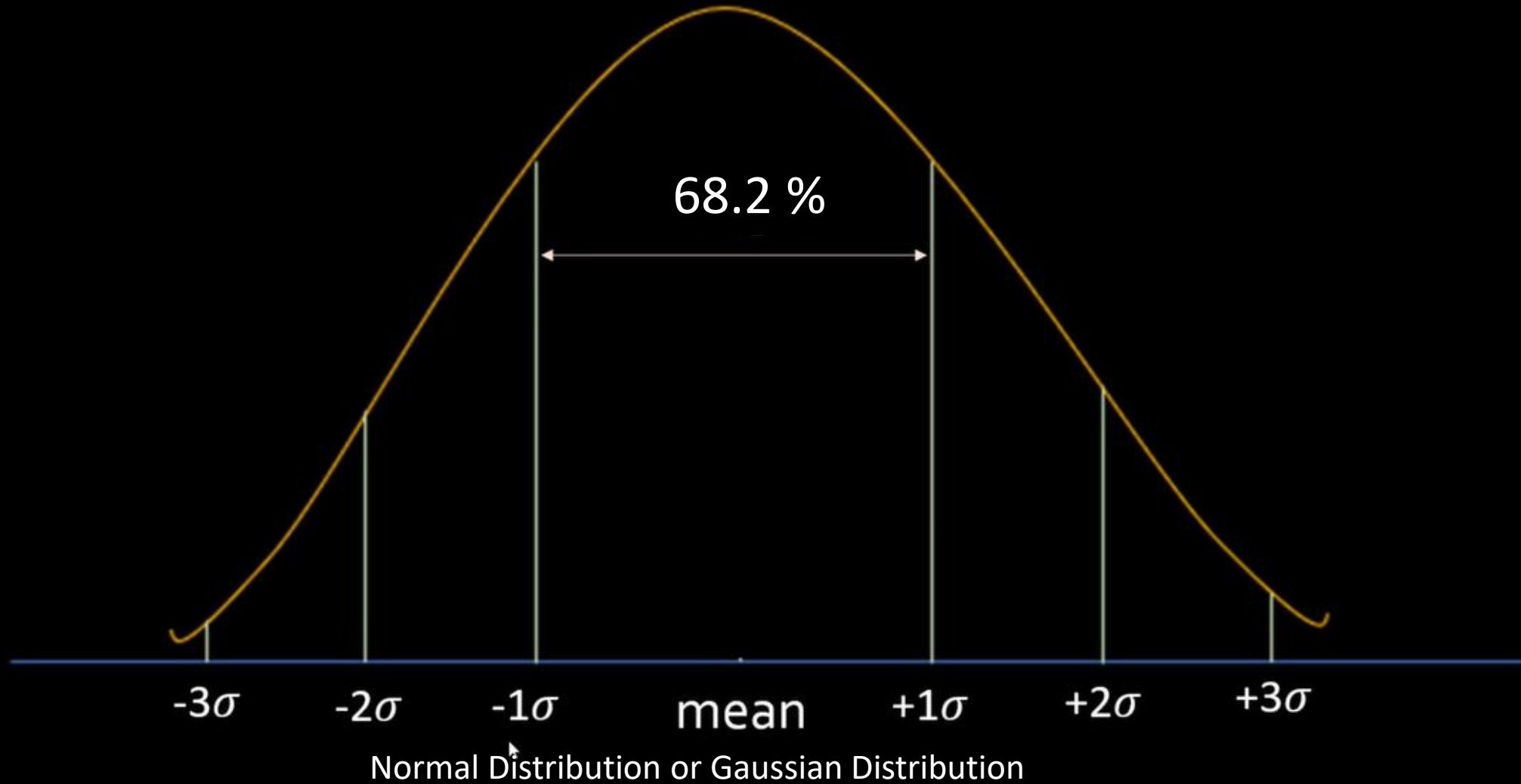
Histogram

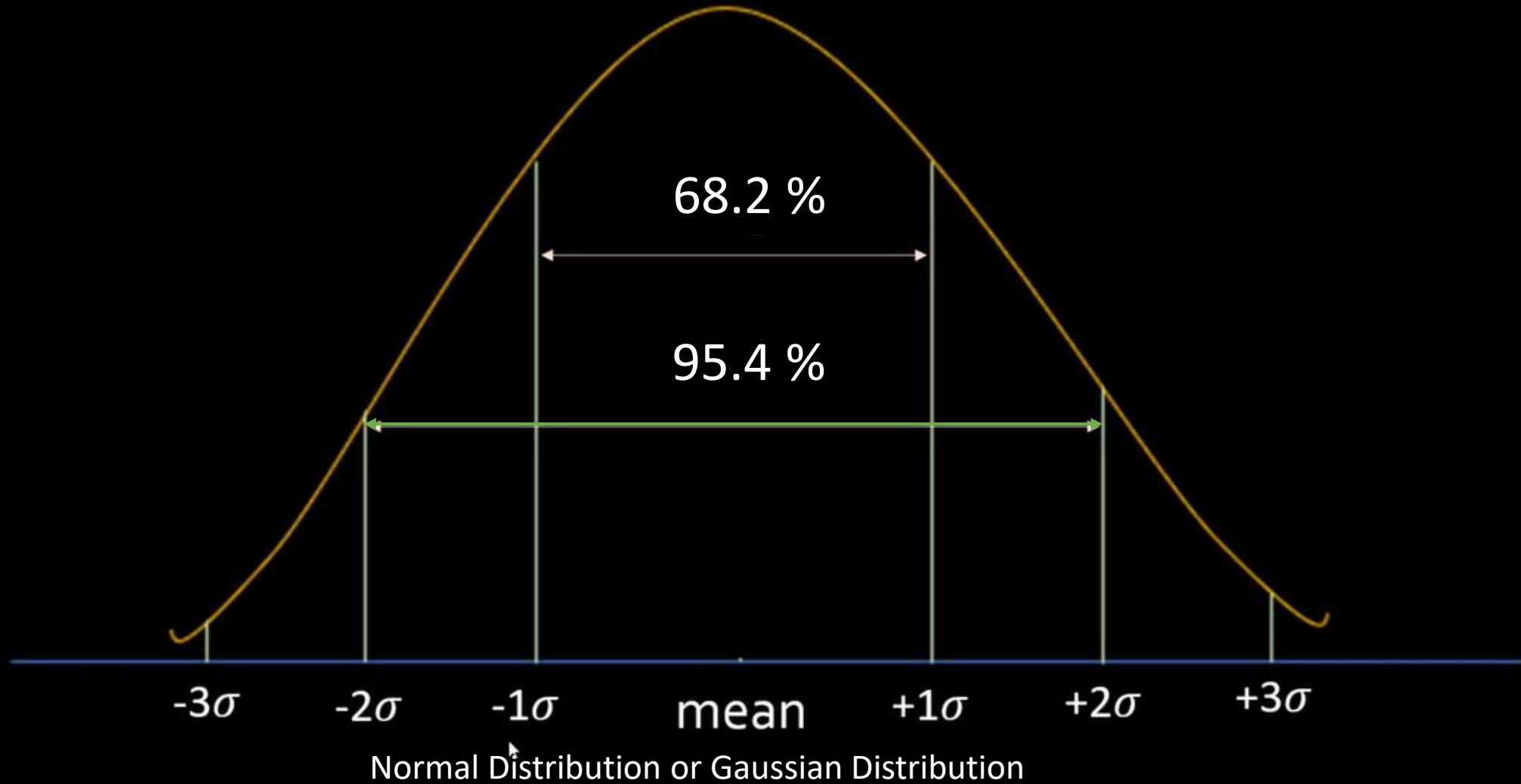


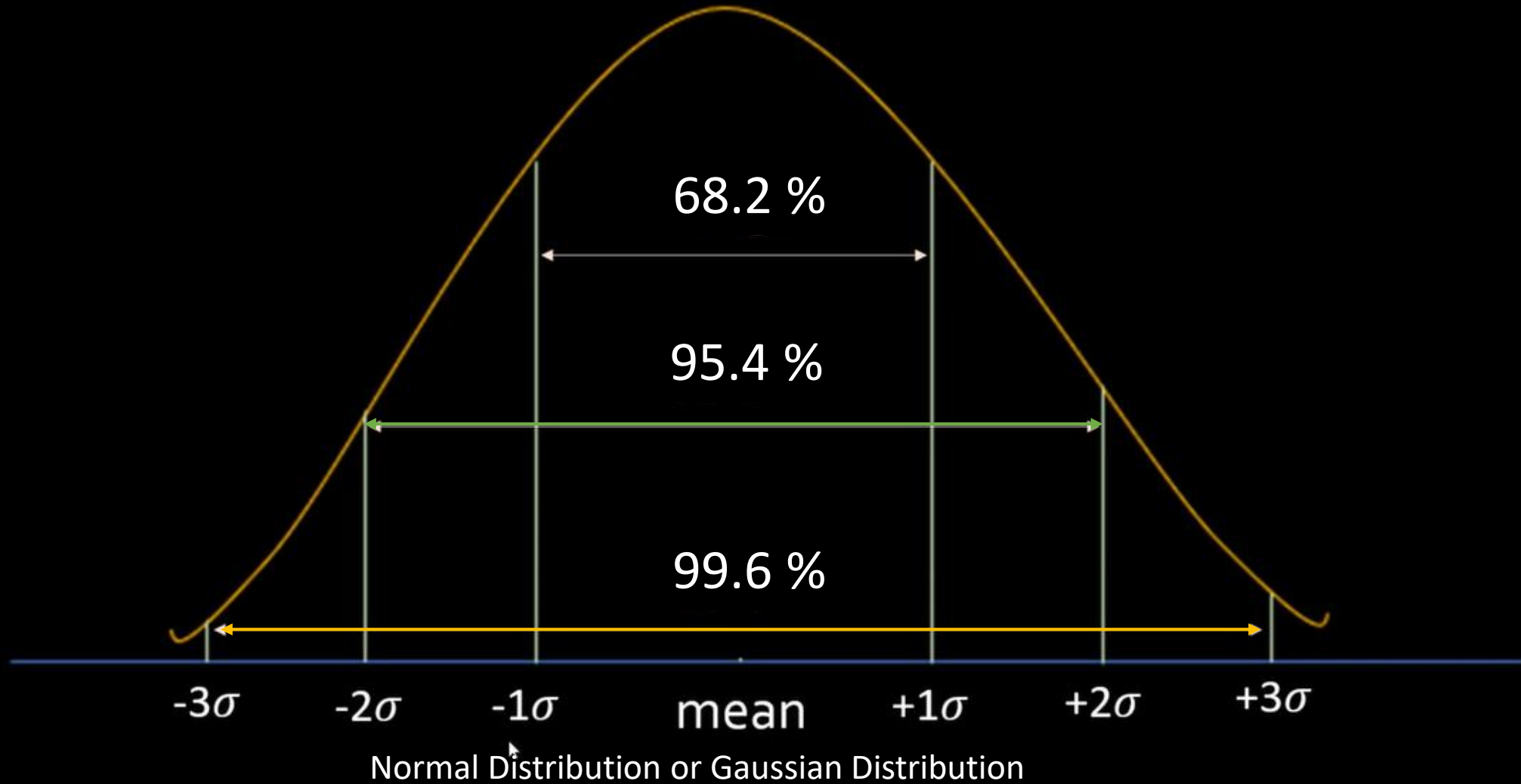
Normal Distribution

- Income Distribution In Economy
- Shoe Size
- Birth Weight
- Spending Days in Hospital









Z Scores

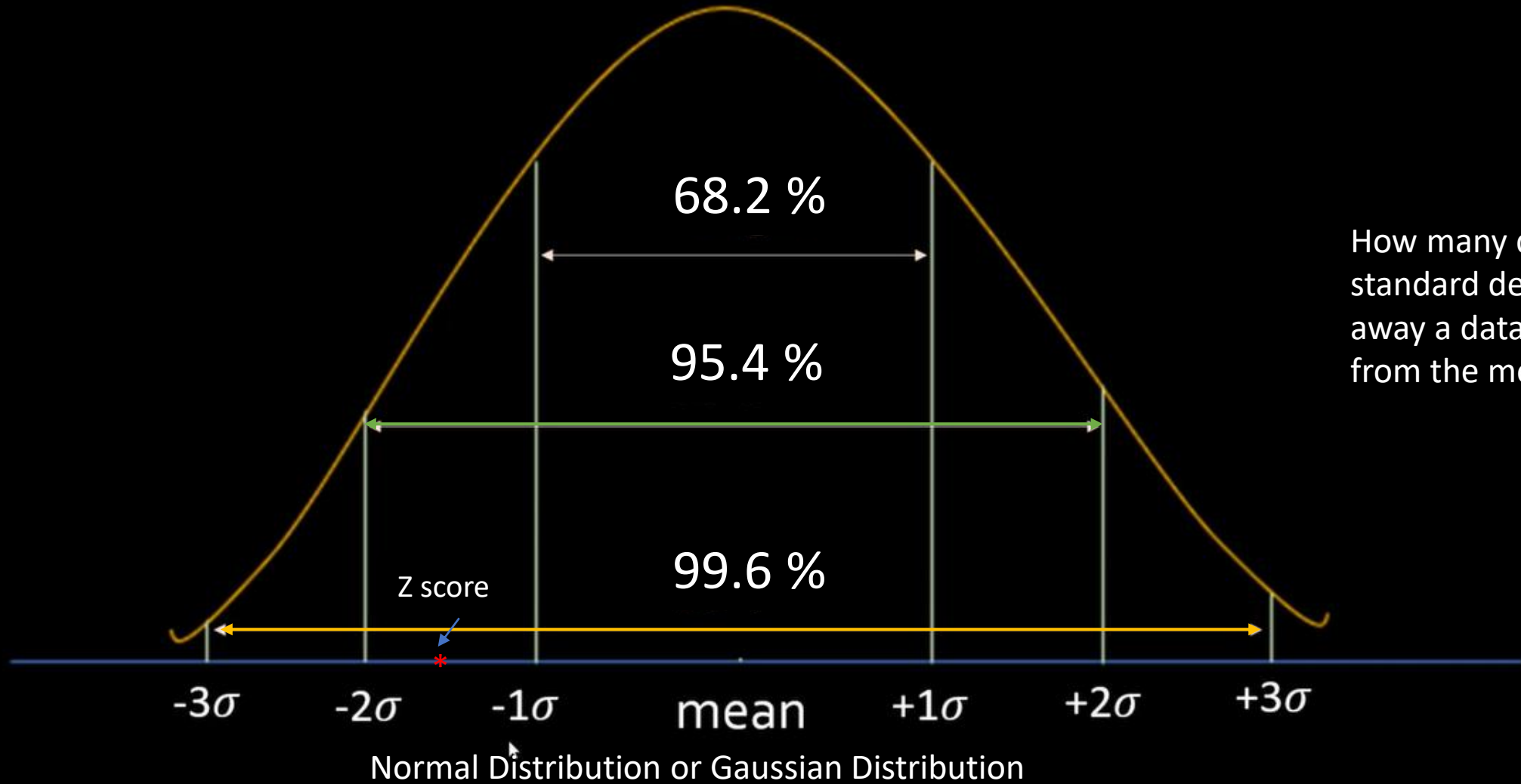
How many distances standard deviation away a datapoint is, from the mean value.

How to Calculate a Z-Score,

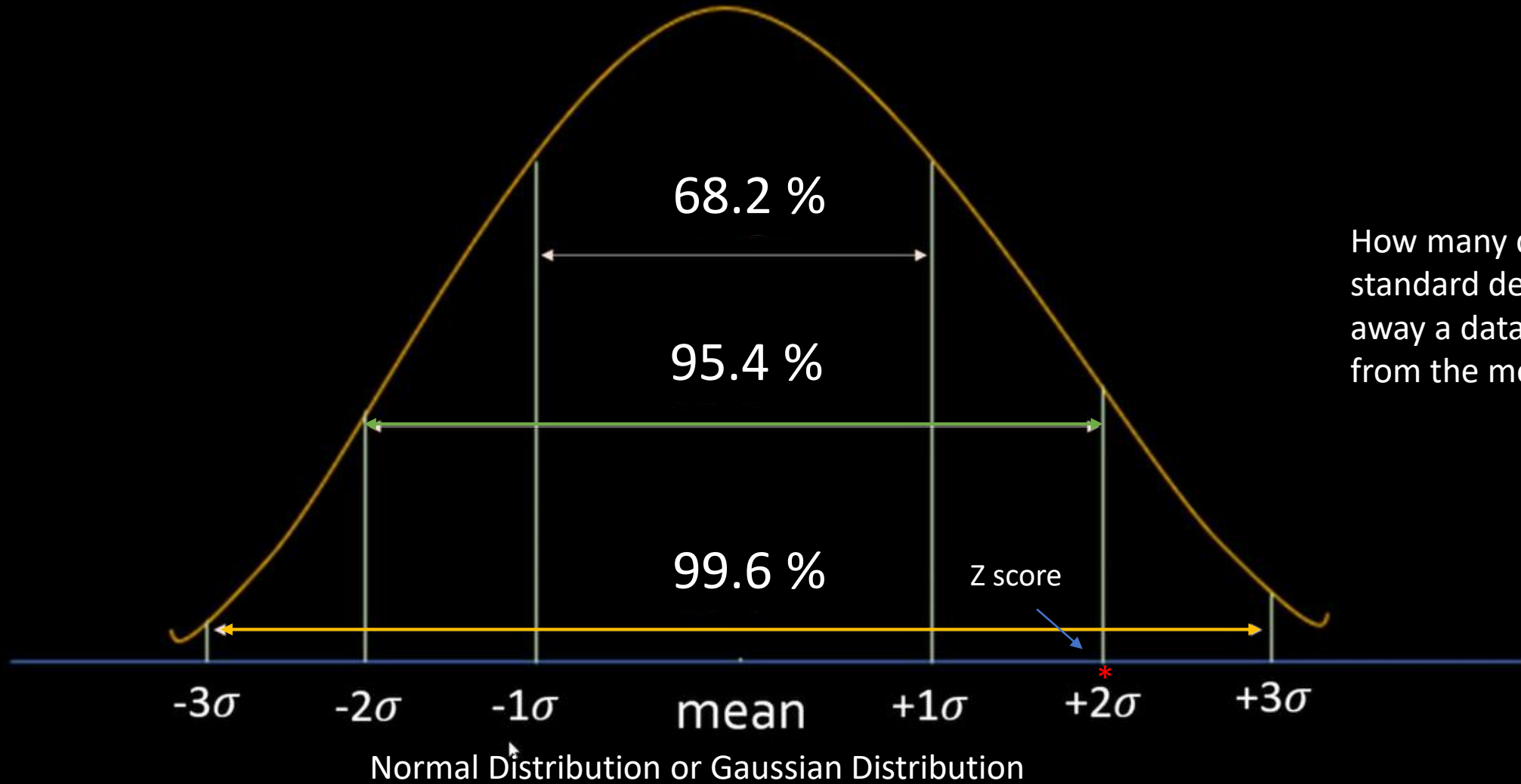
$$Z = \frac{x - \mu}{\sigma}$$

Diagram illustrating the Z-score formula with annotations:

- Score**: Points to x (the datapoint).
- Mean**: Points to μ (the mean value).
- SD**: Points to σ (the standard deviation).



How many distances standard deviation away a datapoint is, from the mean value.



How many distances standard deviation away a datapoint is, from the mean value.