Z SCORE IN GRAPH

This Python function, stdNBgraph(dataset), is designed to **standardize a dataset** and then **visualize the distribution of the resulting Z-scores** using a histogram and kernel density estimate.

Here is a breakdown of the code's steps:

1. Calculating Mean and Standard Deviation

Python

mean = dataset.mean()

std = dataset.std()

- dataset.mean(): Calculates the **arithmetic mean** (average) of all the values in the input dataset. This value () is the center of the original distribution.
- dataset.std(): Calculates the **standard deviation** of the values in the input dataset. This value () measures the amount of variation or dispersion from the mean.

---

2. Preparing Data and Calculating Z-Scores

Python

values = [i for i in dataset]

z_score = [((j-mean)/std) for j in values]

- values = [i for i in dataset]: This line simply converts the input dataset (likely a pandas Series or similar iterable) into a standard **Python list**.
- z_score = [((j-mean)/std) for j in values]: This is the **core standardization step**. It uses a list comprehension to calculate the **Z-score** for every data point () in the values list.

The **Z-score formula** is:

where:

  - is the individual data point ().
  - is the mean of the dataset.
  - is the standard deviation of the dataset.

A Z-score tells you how many standard deviations a data point is away from the mean. The resulting list, z_score, represents the **standardized version** of the original data.

---

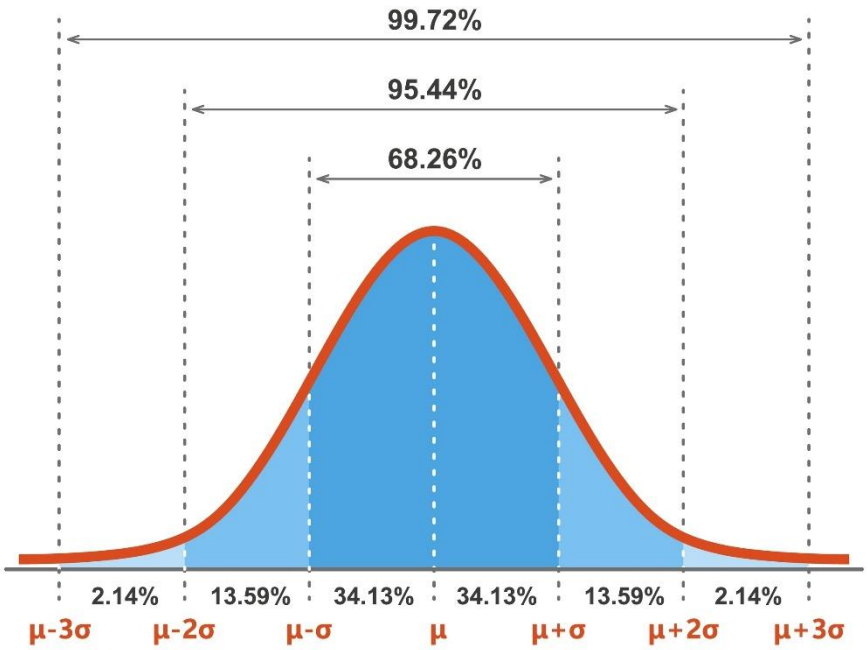
3. Visualization

Python

sns.distplot(z_score, kde=True)

- sns.distplot(...): This function, likely from the **Seaborn** statistical data visualization library (aliased as sns), is used to **plot a distribution**.
- z_score: This is the data being plotted—the list of standardized Z-scores.
- kde=True: This argument instructs distplot to include a **Kernel Density Estimate** line over the histogram, providing a smooth estimate of the probability density function.

Goal: This step visualizes the distribution of the standardized data. If the original data was normally distributed, the plot of the Z-scores will resemble the **Standard Normal Distribution**



---

4. Redundant Calculation

Python

sum(z_score)/len(z_score)

- This line calculates the **mean of the Z-scores**.
- The mean of any standardized dataset (where you subtract the mean and divide by the standard deviation) will theoretically be **0**.
- **Crucially, this result is calculated but not stored or printed, making the line functionally redundant** for the output of the function, though it may have been included to confirm the standardization process.

Summary of Function Purpose

The function stdNBgraph essentially performs **Standardization** (also known as Z-score normalization) on an input dataset, converting it into a new dataset with a **mean of 0** and a **standard deviation of 1**. It then **plots the distribution** of this standardized data.