



SNS COLLEGE OF ENGINEERING



Kurumbapalayam(Po), Coimbatore – 641 107

Accredited by NAAC-UGC with 'A' Grade

Approved by AICTE, Recognized by UGC & Affiliated to Anna University, Chennai

Department of Information Technology

Course Name – 23ADT202 Fundamental of Data
science and Analytics

II Year / IV Semester

Unit 2 – Descriptive Analytics

Z-Scores





Introduction to Z-Scores



Z-scores measure how many standard deviations a data point is from the mean. It helps standardize values, making it easier to compare data across different distributions. Z-scores are essential in identifying outliers and conducting probability calculations. Understanding this concept is key in data analysis.



Formula for Z-Score



The Z-score is calculated using the formula:

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

Where:

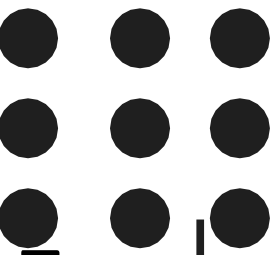
X is the value being analyzed

μ is the mean of the dataset

σ is the standard deviation This formula helps determine how far a data point is from the mean in terms of standard deviations.



Interpretation of Z-Scores



$Z = 0$: The data point is equal to the mean. $Z > 0$: The data point is above the mean. $Z < 0$: The data point is below the mean. The magnitude of Z indicates the extent of deviation; values larger than $|Z| > 2$ suggest significant differences from the mean.



Significance of Z-Scores



Z-scores provide standardization, allowing for easy comparison of values across different datasets or distributions. They also help in detecting outliers—values that significantly differ from the rest of the data. Z-scores further assist in making probability calculations in a normal distribution.



Example Calculation



Consider a dataset: [50, 60, 70, 80, 90]

- Mean (μ) = 70

Standard Deviation (σ) = 15.81 For $X = 90$, the Z-score calculation is:

$$Z = \frac{90 - 70}{15.81} = 1.26$$

This means that 90 is 1.26 standard deviations above the mean.



Identifying Outliers Using Z-Scores

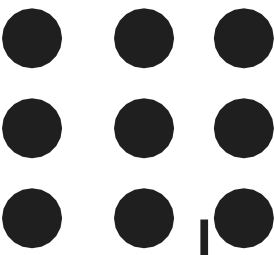


To identify outliers:

1. Compute the mean and standard deviation.
2. Calculate Z-scores for each data point.
3. Identify values where $|Z| > 3$, as they are considered extreme outliers. Outliers can significantly skew analysis, making Z-scores essential in identifying them.



Applications of Z-Scores in Different Fields



Education: Z-scores allow for comparing student performance across different tests

Business: Z-scores help detect outliers in sales data, identifying exceptional performance.

Healthcare: Abnormal lab test results can be flagged using Z-scores, indicating potential health issues.



Exercise: Calculate Z-Score



1. Calculate the Z-score for a data point $X=120$ in a dataset with:

Mean $\mu=100$

Standard Deviation $\sigma=10$ Formula:

$$Z = \frac{120 - 100}{10} = 2$$

The Z-score is 2, meaning the data point is 2 standard deviations above the mean.



Visualizing Z-Scores



Z-scores can be represented on a standard normal curve to visualize their position relative to the mean. A Z-score of 2.5 would lie to the right of the mean, indicating it's more than two standard deviations above the mean. This visualization helps contextualize how extreme or typical a data point is within a distribution.



THANK YOU