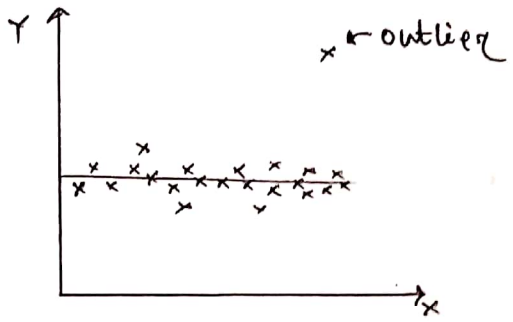
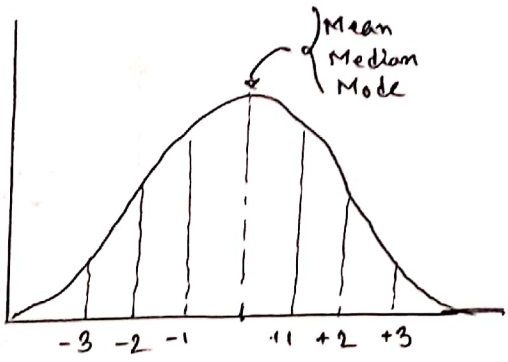
Outlier





Standard Deviation describes how far the data is from Mean.

Types of Outlier Detection Technique:

(1) 3 standard Deviation: Figure out Higher limit and Lower limit based on 3* standard deviation and consider as outlier the points we get outside those limits.

(2) Z score: calculate z score

$$Z = \frac{x - \mu}{\sigma} \quad [\because \mu = \text{mean}, \sigma = \text{standard deviation}]$$

Here, $-3 < z < 3$. Points outside this range is consider as outlier.

Note: 3 is industry convention.

(3) Percentile: Percentile is a measure indicating the value below which a given percentage of observations in a group of observations falls.

Example:

name	test score (from 100)	Percentage	Percentile rank
a	69	69%	100.0%
b	56	56%	50%
c	45	45%	25%
d	32	32%	12%
e	27	27%	0%
f	65	65%	75%
g	61	61%	62%
h	66	66%	87%
i	45	45%	25%

← it means 100% or all data below this value.

← it means there is no data below this value.