

Types of Data

Record: data matrix (crosstabs), document data (term-frequency vector/text documents)

Graph/Network: WWW, facebook, molecular structures

Ordered: Video data (sequence of images), temporal data - time-series, genetic sequence data

Spatial/Image/Multimedia: Maps, Photos, Videos

Median interval

$$median = L_1 + \left(\frac{n/2 - (\sum freq)_i}{freq_{median}} \right) width$$

Median is difficult to calculate for large amounts of data, so approximated/interpolated for grouped data to median interval. L_1 is lower boundary of mdn interval, N is # of vals of entire dataset, $freq$ is the sum of $freq$ of all lower than mdn interval, $freq_median$ is $freq$ of mdn interval, and $width$ is the width of mdn interval.

Attribute Type: Just important info

Binary attribute type?

Under nominal attribute type: categories subtype and also discrete

Symmetric binary vs asymmetric binary

Outcomes equally important vs not eqly important

Numeric: interval-scaled vs ratio-scaled

No true 0 pt, > *temperature, not in kelvin* True 0 pt, ratios > : *temperature kelvin, length, count*

Measures of central tendency: Mode/Midrange

Unimodal, multimodal, bimodal, trimodal, no mode

Datasets with one mode vs more than one mode vs two modes vs 3 modes vs each val only once

unimodal data formula

assymetrical, formula: $mean - mode = 3 * (mean - median)$

symmetric vs positively vs negatively skewed data

mean=median=mode @ same center vs mode < median < mean (right-skewed) vs mean < median < mode

midrange

highest+lowest_val divided by 2

Measures of central tendency: Mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \mu = \frac{\sum x}{N}$$

$$\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$$

1st one is sample mean, 2nd is population mean, 3rd is weighted mean.
 👍 Most useful measure of center 👎 Bad for skewed/outliers
 Solution: trimmed mean: mean after trimming outliers. 👎 Loss of valuable info if too much trimmed down.



By ntp
cheatography.com/ntp/

Not published yet.
Last updated 12th December, 2014.
Page 1 of 1.

Sponsored by **Readability-Score.com**
Measure your website readability!
<https://readability-score.com>