Descriptive Statistics

NASHVILLE SOFTWARE SCHOOL

Descriptive Statistics - Overview

- Used to summarize the characteristics of a dataset
- Important part of the data exploration process
- Some descriptive statistics functions available in spreadsheets:
  - Measures of central tendency
  - Measures of variability
  - Percentile and quartile functions

Descriptors of Central Tendency - Mean

SUM(array)/COUNT(array)

Mean – Often generically called the “average”. Among the most easily recognizable measures of central tendency thus able to be used when communicating to people with minimal statistics/analytics experience.

Pros:

- Weighs all elements of an array equally

Cons:

- Can be overly influenced by extreme values and outliers

Descriptors of Central Tendency - Median

The value of the nth item of a sorted array where n = LENGTH(array)/2

Median – Is a measure that describes the center of a sorted group. It is both recognizable and easily described, this makes it a good tool for communicating.

Pros:

- Not as affected by outliers

Cons:

- Data must be sorted
- May not detect meaningful fluctuations within the data

Descriptors of Central Tendency - Mode

The maximum frequency of all values in an array

Mode – A measure used to describe the value(s) which occur most frequently. Most likely to match the value of an item chosen at random from an array.

Pros:

- Useful at identifying peak values
- If data is sorted can identify transitions between peaks and valleys of frequency

Cons:

- May not be useful if data is continuous
- May not be useful even if data is continuous

Percentile and Quartiles

Percentile – The value in which the given percentage of data falls below when sorted.

Quartiles – Specific, equally-spaced, percentiles that divide the data into 4 equal parts.