Statistics Basics to Advanced

Friday, October 7, 2022 11:55 PM

Central Tendency:

1. Mean
2. Median
3. Mode

Refers to the measure used to determine the center of the distribution of data.

Outliers: Live Day 2- Basic To Intermediate Statistics

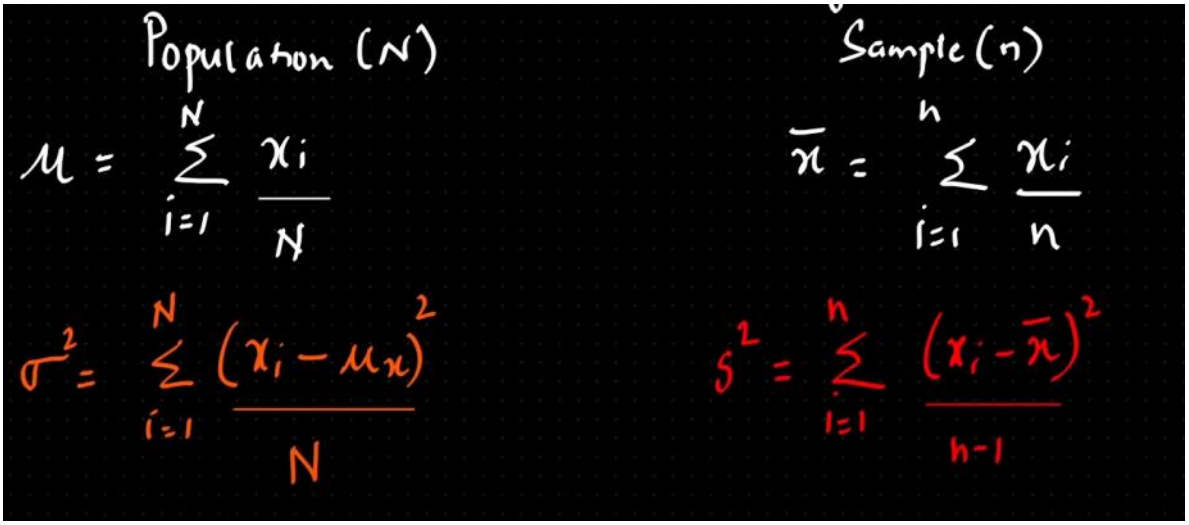
Use case of Mean, Median and Mode:
Live Day 2- Basic To Intermediate Statistics

If there are some missing data, then mode is used so that missing data can be replaced with the most frequent occurring element; this works best for categorical/Qualitative data.

In case of numerical data, use of mean is most preferred.(*see video plzz)

Measure of Dispersion:

1. Variance
2. Standard Deviation



Variance:

Live Day 2- Basic To Intermediate Statistics

Higher the variance,

greater will be the dispersion of data

More will be the graph spreaded

Percentiles and Quartiles:

Live Day 2- Basic To Intermediate Statistics

Percentile is a value below which a certain percentage of observation lie.

Percentile Rank of x = (# of values below x)/n * 100
Index Position/Value Position = (Percentile/100)*(n+1)

Five Number Summary:

1. Minimum
2. First Quartile (Q1) => 25%ile
3. Median
4. Third Quartile (Q3) => 75%ile
5. Maximum

By the help of FNS, we remove outliers.

Removing Outliers and box plot

Lower Fence: Q1 - 1.5(IQR)
Upper Fence: Q3 + 1.5(IQR)
Interquartile Range (IQR) = Q3 - Q1

Why sample variance is divided by n-1 ??