# Basic to Intermediate

1 Measuring of Central Tendency

2 Measure of dispersion

3 Gaussian Distribution

4 Z score

5 Standard Normal Distribution

1 Arithmetic mean for population and sample

Mean (Average) – Trying to find average of specific distribution.

Population (N) Sample (n)

X = {1,2, 3,4,5,6, }

µ = or X’/N X’= = x’/n =

µ =1+1+2+2+3+3+4+5+5+6/N x’ = 3.2

µ=3.2

1 Central Tendency

1 Mean 2 Median 3 Mode

Mean - Refers to the measure used to determine the centre of the distribution of data.

X= 1,1,2,2,3,4,5,5,6 (100) outliers

Mean = 3.2

If we add one more item 100.

Then mean will be 32+100/11 =132/11 = 12.

Vast different in mean because of outliers.

Median {1,1,2,2,3,3,4,5,5,6,100}

1 Sort the numbers

2 Central element

Odd number =11; Central element 3

Median =3

If even data set {1,1,2,2,3,3,5,5,6,100,112}

3+4 = 7/2 = 3.5

3 Median = 3 No outliers

Media = 3.5 with outliers

Note Median works well with outlier

3 Mode

X= {1,2,2,3,4,5,6,6,6,7,8,100,200}

Mode = {Most frequent element}

Maximum repeated variable 6 3 times which also reassure of central tendency.

Disadvantage – If there are many outliers.

{1,2,2,3,4,5,6,6,6,7,8,100,100,100,100}

Mode = will be 100, Which is not true, in this case median is best.

Where mode is considerably use. e.g

Mode / In this data set 10% missing.

Type of flower Petal length Petal width

Rose

Lilly Missing Value will be replaced Most frequent occurring element.

Sunflower this works well with categorical variable

---

--

--

Example.

Age of Study? In this case which central tendency should be apply

25 Mean, Mode, Median,

26 If this is the case of world population, we will not use mean.

-

-

-

32

34

38

Measure of Dispersion (Dispersion means spread) ( How well spread your dat).

1 Variance

2 Standard Deviation

1 Variance - Variance is concept measuring of Dispersion.

Data set 1 Set 2

X’ = {1,1,2,2,4} x’= {2,2,2,2,2}

X’=2 x’=2

In above example both distributions have some average(mean) -2

That means this mean doesn’t identify these distributions are different.

How two distributions are different!



Population variance Sample variance

62 =(xi-u)2/N =10.8/6 =1.81 s2 =(xi-x’)2/n-1

X u x-u (x-u)2

1 2.83 -1.83 3.34

2 2.83 -0.83 0.6889

2 2.83 0.83 0.6889

3 2.83 0.17 0.03

4 2.83 1.17 1.37

5 2.83 2.17 4.71

U=2.83 10.84 =U 10.84/6 =1.81

6 = = 1.345

1 U=1.81 where variance is more 2 U= 5.45

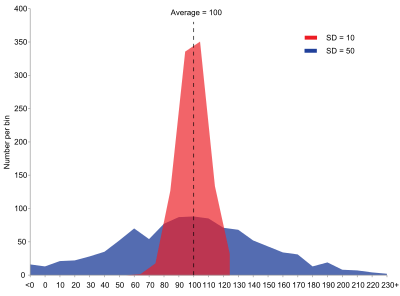
Low high

If distribution diagram 1 has less variance than 2. High mean high variance. Less mean less variance.

Or element present in central region is more.

U =5.45

https://en.wikipedia.org/wiki/Variance



Eg

In above example, Same mean from two population but different variance, The red population has mean 100 and variance 100 (SD=10) while other population has means 100 and variance 2500 (SD=50).

0.15 1.49 4.17 u=2.83 4.17 5.51

2.83 2.83 4.17 1.49

1.34 1.34 1.34 1.34

---- ----- ------- ------

4.17 1.49 5.51 0.15

With variance we can figure out data spread

SD between one standard deviation to the right and left what may be the range of data that may be falling.

Eg in above example

How far element can be, let take 5

# How you represent the 5. = It falls 1.5 SD from the mean