Ms. Shweta M. Pandya

Assistant Professor

Dept. of Pharmacology

Saraswati Institute of Pharmaceutical Sciences,Dhanap, Gandhinagar- 382355

Lecture.5

Measuresofdispersion-Range,Variance-Standarddeviation–co-efficientof variation - computation of the above statistics for raw and grouped data

MeasuresofDispersion

The averages are representatives of a frequency distribution. But they fail to givea complete picture of the distribution. They do not tell anything about the scatterness of observations within the distribution.

Suppose that we have the distribution of the yields (kg per plot) of two paddy varieties from 5 plots each. The distribution may be as follows

Variety I	45	42	42	41	40
Variety II	54	48	42	33	30

It can be seen that the mean yield for both varieties is 42 kg but cannot say thatthe performances of the two varieties are same. There is greater uniformity of yields inthe first variety whereas there is more variability in the yields of the second variety. The first variety may be preferred since it is more consistent in yield performance.

Form the above example it is obvious that a measure of central tendency alone is not sufficient to describe a frequency distribution. In addition to it we should have a measure of scatterness of observations. The scatterness or variation of observations from their average are called the dispersion. There are different measures of dispersion like the range, the quartile deviation, the mean deviation and the standard deviation.

Characteristicsofagoodmeasureof dispersion

Anidealmeasure ofdispersionisexpectedtopossessthefollowing properties

1. It should be rigidly defined
2. It should be based on all the items.
3. It should not be unduly affected by extreme items.
4. It should lend itself for algebraic manipulation.
5. It should be simple to understand and easy tocalculate

Range

Thisisthesimplestpossiblemeasureofdispersionandisdefinedasthedifference between the largest and smallest values of the variable.

- In symbols, Range = L– S.
- Where L = Largest value.
- S= Smallest value.

Inindividualobservationsanddiscreteseries,LandSareeasilyidentified. In continuous series, the following two methods are followed.

Method1

L=Upperboundaryofthehighestclass S= Lower boundary of the lowest class.

Method2

L=Midvalueofthehighestclass. S = Mid value of the lowest class.

Example1

Theyields(kgperplot)ofacottonvarietyfromfiveplotsare8,9,8,10and11.Findthe range

Solution

L=11, S= 8.

Range = L– S= 11- 8 = 3

Example 2

Calculate range from the following distribution.

Size:	60-63	63-66	66-69	69-72	72-75
Number:	5	18	42	27	8

Solution

L=Upperboundaryofthehighestclass=75 S =

Lower boundary of the lowest class = 60

Range = L – S = 75 – 60 = 15

MeritsandDemeritsofRange

Merits

1. Itissimple to understand.
2. Itiseasy to calculate.
3. Incertaintypesofproblemslikequalitycontrol,weatherforecasts,shareprice analysis, etc., rangeismost widelyused.

Demerits

1. Itisvery much affected by the extreme items.
2. Itisbased on only two extreme observations.
3. Itcannotbecalculatedfromopen-endclassintervals.
4. Itisnot suitable for mathematical treatment.
5. Itisa very rarely used measure.

StandardDeviation

Itisdefinedasthepositivesquare-rootofthearithmeticmeanoftheSquareofthe deviations of the given observation from their arithmetic mean.

Thestandarddeviationisdenotedbysincaseofsampleand Greekletter σ (sigma) in case of population.

Theformula for calculating standarddeviation isasfollows

$$s = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}}$$ for raw data

Andforgroupeddatatheformulasare

$$s = \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2}$$ for discrete data

$$s = C \times \sqrt{\frac{\sum fd^2}{N} - \left(\frac{\sum fd}{N}\right)^2}$$ forcontinuousdata

Whered = $\frac{x - A}{C}$

C=class interval

Example3

Raw Data

Theweightsof5ear-headsofsorghumare100,102,118,124,126gms.Findthestandard deviation.

Solution

x	**x^2**
100	10000
102	10404
118	13924
124	15376
126	15876
570	**65580**

Standard deviation$s = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}}$

$$= \sqrt{\frac{65580 - \frac{(570)^2}{5}}{5-1}} = \sqrt{150} = 12.25 \; gms$$

Example4

Discretedistribution

The frequency distributions of seed yield of 50 seasamum plants are given below.

Findthe standard deviation.

Seedyieldingms(x)	3	4	5	6	7
Frequency (f)	4	6	15	15	10

Solution

Seedyield ingms(x)	**f**	**fx**	**fx^2**
3	4	12	36
4	6	24	96
5	15	75	375
6	15	90	540
7	10	70	490
Total	**50**	**271**	**1537**

Here n = 50

Standarddeviation $s = \sqrt{\frac{\sum fx^2}{n} - \left(\frac{\sum fx}{n}\right)^2}$

$= \sqrt{\frac{1537}{50} - \left(\frac{271}{50}\right)^2}$

$= \sqrt{30.74 - 29.3764}$

= 1.1677 gms

Example5
Continuousdistribution

TheFrequencydistributionsofseedyieldof50seasamumplantsaregivenbelow.Find the standard deviation.

Seedyield ingms(x)	2.5-35	3.5-4.5	4.5-5.5	5.5-6.5	6.5-7.5
No.ofplants(f)	4	6	15	15	10

Solution

Seedyield ingms(x)	**No.of Plants f**	**Mid x**	**d=** $\frac{x-A}{C}$	**df**	**d^2f**
2.5-3.5	4	3	-2	-8	16
3.5-4.5	6	4	-1	-6	6
4.5-5.5	15	5	0	0	0
5.5-6.5	15	6	1	15	15

6.5-7.5	10	7	2	20	40
Total	**50**	**25**	**0**	**21**	**77**

A=Assumed mean = 5

n=50, C=1

$$s = C \times \sqrt{\frac{\sum fd^2}{n} - \left(\frac{\sum fd}{n}\right)^2}$$

$$= 1 \times \sqrt{\frac{77}{50} - \left(\frac{21}{50}\right)^2}$$

$$= \sqrt{1.54 - 0.1764}$$

$$= \sqrt{1.3636} = 1.1677$$

MeritsandDemeritsofStandardDeviation Merits

1. Itisrigidlydefinedanditsvalueisalwaysdefiniteandbasedonalltheobservations and the actual signs of deviations are used.
2. Asitisbasedonarithmetic mean,ithasallthemeritsofarithmetic mean.
3. Itisthemost importantand widelyused measureof dispersion.
4. Itispossible for further algebraic treatment.
5. Itislessaffectedbythefluctuationsofsamplingandhencestable.
6. Itisthebasisformeasuring thecoefficient ofcorrelation and sampling.

Demerits

1. Itisnot easyto understand and itisdifficult to calculate.
2. Itgivesmoreweighttoextremevaluesbecausethevaluesaresquaredup.
3. Asitisanabsolutemeasureofvariability,itcannotbeusedforthepurposeof comparison.

Variance

Thesquare of the standarddeviation iscalled variance (i.e.)variance=$(SD)^2$.

Coefficient of Variation

The Standard deviation is an absolute measure of dispersion. It is expressed in terms of units in which the original figures are collected and stated. The standard deviation of heights of plants cannot be compared with the standard deviation of weights of the grains, as both are expressed in different units, i.e heights in centimeter andweights in kilograms. Therefore the standard deviation must be converted into a relative measure of dispersion for the purpose of comparison. The relative measure is known as the coefficient of variation. The coefficient of variation is obtained by dividing the standarddeviationbythemeanandexpressedinpercentage.Symbolically,Coefficientof variation (C.V) =$\frac{SD}{mean}\times 100$

If we want to compare the variability of two or more series, we can use C.V. The series or groups of data for which the C.V. is greater indicate that the group is more variable,lessstable,lessuniform,lessconsistentorlesshomogeneous.IftheC.V.isless, it indicates that the group is less variable or more stable or more uniform or more consistent or more homogeneous.

Example6

Consider the measurement on yield and plant height of a paddy variety. The mean and standard deviation for yield are 50 kg and 10 kg respectively. The mean and standard deviation for plant height are 55 am and 5 cm respectively.

Here the measurements for yield and plant height are in different units. Hence the variabilities can be compared only by using coefficient of variation.

For yield, CV=$\frac{10}{50}\times 100$= 20%

For plant height, CV=$\frac{5}{55}\times 100$= 9.1%

Theyield issubject to more variation than the plant height.

Questions

1. Whichmeasure isaffected most bythe presence of extreme values.

a) Range b)Standard Deviation

b) Quartile Deviation d) Mean deviation

Ans:StandardDeviation

2. Variance is square of _____________

a) Range b)Standard Deviation

c) QuartileDeviation d) Mean deviation

Ans:StandardDeviation

3. Ifthe CVof varietyI is30% and varietyII is25% thenVariety II ismore consistent.

Ans:True

4. Forthe set of data 5, 5, 5,5,5,5 the Standard deviation value iszero.

Ans:True

5. Theabsolutemeasuresofdispersion willhavetheoriginal units.

Ans:True

6. The mean deviation value for a set of data can take even negative value.

Ans: False

7. Define dispersion.

8. DefineC.V. Whatare itsuses?

9. Whatarethedifferencesbetweenabsolutemeasureandrelativemeasureof dispersion?

10. Howto calculatethe standard deviationfor rawand grouped data?