MEASURES OF DISPERSION

RANGE, QUARTILE DEVIATION

சிதறல் அளவைகள்

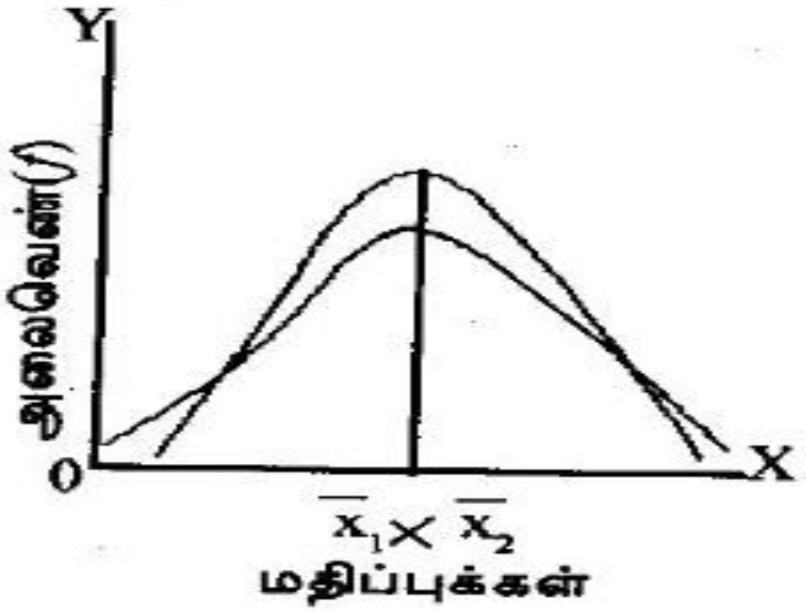
வீச்சு , கால்மான விலக்கம்

DR.S.ARULJOTHISELVI
ASSISTANT PROFESSOR
DEPARTMENT OF ZOOLOGY
PERIYAR GOVERNMENT ARTS COLLEGE
15.09.2020

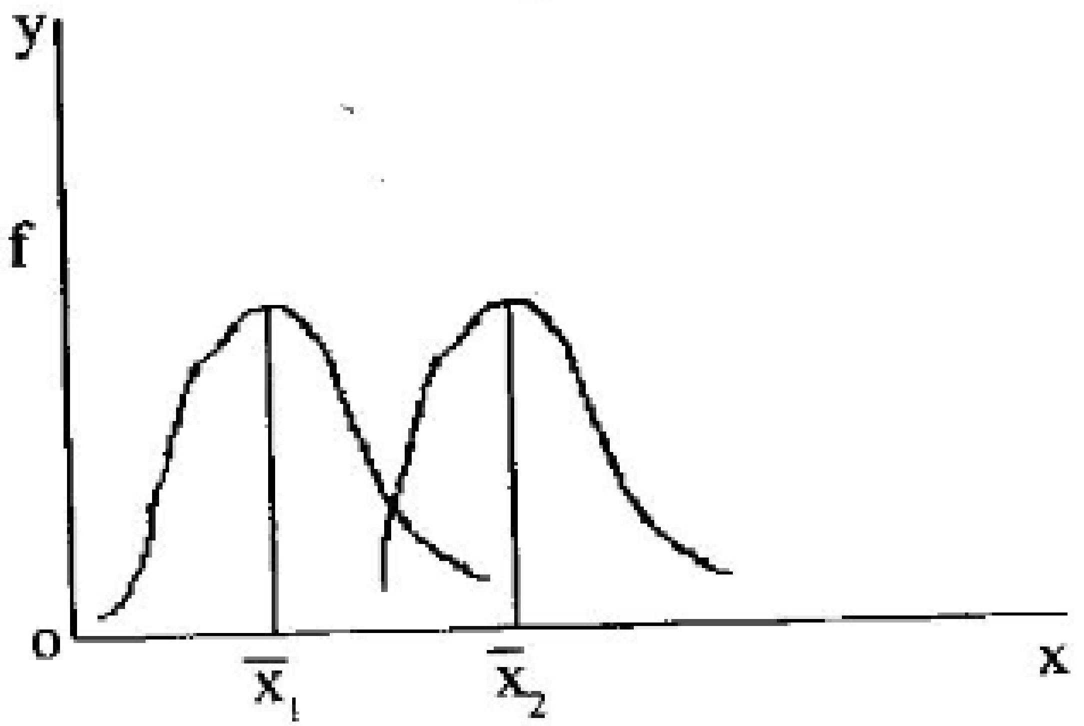
சிதறல் அளவைகள்

(MEASURES OF DISPERSION)

இதுவரை ஒரு பரவலுக்கு மையநிலைப் போக்கு அளவைகளைக் கண்டோம். சில வேளைகளில் மைய நிலைப்போக்கு அளவைகள் பரவலின் தன்மையைச் சிறப்பாக விளக்குவதில்லை. இரண்டு பரவல்களின் கூட்டுச்சராசரி ஒன்றாக இருக்கலாம். அவற்றின் அலைவெண்கள் சமமாக இருககலாம். ஆனாலும் இரு பரவல்களும் ஒத்தவை எனக்கொள்ளமுடியாது. ஒரு பரவலில் (பரவல் -A)



மற்றமதிப்புக்கள் சராசரியை ஒட்டி அமையலாம். மற்ற பரவலில (பரவல் B) சராசரியை விட்டு மற்ற மதிப்புக்கள் விலகி இருக்கலாம். எனவே தனிமதிப்புக்கள் ஒரு சராசரியிலிருந்து எவ்வளவுக்கு விலகி உள்ளன என்பதனை அறிய வேண்டியுள்ளது. இன்னும் இரு பரவல்களை எடுத்துக் கொள்வோம். இவற்றின் சராசரிகள் வேறு வேறு. ஆனால் அவற்றில் மதிப்புக்கள் சராசரியைச் சுற்றி ஒரே மாதிரியாக அமைந்துள்ளன. கீழ்கண்ட படம் இதனை விளக்கும்.



சிதறல் அளவை காண்பதன் நோக்கங்கள

ஒரு சராசரியைச் சுற்றி தனிமதிப்புக்கள் எவ்வாறு அமைந்துள்ளன என்பதைக் காண்பதே சிதறல் அளவை அல்லது பரவு வகை அளவையின் நோக்கமாகும் (Measures of dispersion or scatter).

1. பின்வரும் உதாரணத்தைக் காண்போம். ஒரு மாணவன் 3 பாடங்களில் பெற்ற மதிப்பெண்கள் பின்வருமாறு.

பொருளியல்	50	50	50	50	50	50
புள்ளியியல்	60	70	30	40	55	45
சரித்திரம்	80	20	90	10	75	25

மூன்று பாடங்களிலும் சரியாக 50% வாங்கியுள்ளான். இருப்பினும் பொருளியலில் நிலைப்புத்தன்மை அதிகமாகவும், புள்ளியியலில் நிலைப்புத்தன்மை (Consistency) குறைவாகவும், சரித்திரத்தில் நிலைப்புத்தன்மையற்றும் காணப்படுவதை அறியலாம்.

இதனால் ஒரு சராசரி எவ்வளவு தூரம் நம்பகமானது என்பதை அறிய இது பயன்படுவதைக் காண்கிறோம். சிதறல் அதிகமிருந்தால் நிலைப்புத்தன்மை குறைவு என அறிகிறோம்.

காவேரி ஆற்றில் வெள்ளம் பாய்கிறது. சராசரி ஆழம் 5 அடி 6 அங்குலம். நமது உயரம் 5 அடி 10 அங்குலம். எனவே ஆற்றை நடந்து கடந்துவிடலாம் என்று எண்ணி ஆற்றில் இறங்கலாமா? சில இடங்களில் ஆழம் 7 அடி இருக்கலாம். நீச்சல் தெரியாதவரின் கதி என்னவாகும்? எனவே சராசரியைவிட சிதறல் அளவை நமக்கு சராசரியில் இருந்து மற்ற மதிப்புக்கள் சராசரியாக எவ்வளவு தூரம் விலகியுள்ளன என்பதை அறிய உதவும்.

2. மாறுபாட்டினை (Variability) அறிந்து அதனைக் குறைக்க நடவடிக்கை எடுக்கலாம். ஒரு நோயாளியின் உடல் சூடு இரத்த அழுத்தம் ஏறி இறங்குவதை அறிந்து அதனைக் கட்டுப்படுத்தலாம்.

3. ஒப்பிட்டு நோக்கித் திறமைசாலிகள் யார் என்பதைத் தேர்ந்தெடுக்கலாம். ஒரு கிரிக்கெட் வீரர் எந்த அளவு நிலைப்புத்தன்மையுடன் ஓட்டங்கள் எடுக்கிறார் என்ற கண்டுகொள்ள சிதறல் அளவை பயன்படும். அதிக நிலைப்புத்தன்மை உடையவரே சிறந்த ஆட்டக்காரர் ஆவார். இரு கிரிக்கெட் வீரர்கள் 10 ஆட்டங்களில் எடுத்த ஓட்டங்கள் பின்வருமாறு:

A	75	35	25	45	30	80	10	50	140	55	90	75
B	75	70	60	8	65	75	90	80	60	85	80	80

இதில் இரண்டாவது (B) ஆட்டக்காரர் முதல் ஆட்டக்காரரை (A). விட அதிக நிலைப்புத்தன்மையுடையவர் என அறியலாம்.

4. சிதறல் அளவையைப் பயன்படுத்தி செல்வம். வருவாய் ஆகிய பகிர்வின் உண்மைநிலையை அறியலாம். சமப்பகிர்வா, இல்லையெனில் எவ்வளவு தனி வருவாய்கள் சிதறி உள்ளன என்பதை அறியலாம். இரு பரவல்களின் சிதறல் தன்மையை ஒப்பிட்டு நோக்கலாம்.

5. தொழில் உற்பத்தியில் தரக்கட்டுப்பாட்டிற்கு இந்த அளவைகள் பெரிதும் பயன்படுகின்றன. பொருளின் தரத்தைக் கூட்டி விற்பனையைப் பெருக்கலாம்.

6. சிதறல் அளவைகள் மென்மேலும் பரவலின் பல தன்மைகளை அறியப் பயன்படுகின்றன (உ - ம்) மாறு விகிதக்கெழு, கோட்டக்கெழு.

சிதறல் அளவையின் வகைகள்

(1) வீச்சு (Range)

இது மிக எளிமையான சிதறல் அளவை ஆகும். மிக அதிக மதிப்புக்கும் மிகக் குறைந்த மதிப்புக்கும் உள்ள வேறுபாடே வீச்சாகும். அலைவெண் வளைகோட்டின் இரு முனைகளுக்கிடையே உள்ள தூரமே வீச்சு ஆகும். அலைவெண் பலகோணமாயின் முதற்பிரிவு இடைவெளியின் கீழ் எல்லைக்கும் கடைசிப் பிரிவு இடைவெளியின் மேல் எல்லைக்கும் இடையே உள்ள தூரம் வீச்சு ஆகும்.

வீச்சு = L - S ;	வீச்சுக்கெழு $= \dfrac{L - S}{L + S}$

L = மிகப்பெரிய மதிப்பு, S = மிகச்சிறிய மதிப்பு

மாதிரி: 1 பின்வரும் விவரங்களிலிருந்து வீச்சு வீச்சுக்கெழு காண்க.

பங்குவிலை: 100, 150, 50, 200, 250

வீச்சு = 250 − 50 = 200

$$\text{வீச்சுக்கெழு} = \frac{250 - 50}{250 + 50} = \frac{200}{300} = 0.67$$

மாதிரி: 1. பின்வரும் அலைவெண் பரவலில் இருந்து வீச்சு காண்க.

பிரிவு இடைவெளி	அலை வெண்
50-60	5
60-70	15
70-80	20
80-90	25
90-100	5

வீச்சு = L - S = 100 − 50 = 50

$$\text{வீச்சுக்கெழு} = \frac{L - S}{L + S}$$

$$= \frac{100 - 50}{100 + 50} = \frac{50}{150} = 0.33$$

நிறைகள்:

1. வீச்சு மிக எளிமையானது; எளிதில் கணக்கிடலாம்.

2. விரைவில் காணக்கூடிய அளவையாகும். சிலவேளை களில் உடனடியாகச் சிதறல் தன்மையைக் காணவேண்டி இருந்தால் இதனைப் பயன்படுத்தலாம்.

3. தரக்கட்டுப்பாட்டிலும், பங்குச்சந்தை விலை வீச்சு காணவும்; நோயாளியின் உடல் சூடு, இரத்த அழுத்தம் முதலியவற்றை அறிந்து நடவடிக்கை எடுக்கவும், பருவகால நிலைகளை ஆராயவும் இது பயன்படுகிறது.

குறைகள்:

1. வீச்சு எல்லா எண்களின் அடிப்படையில் கணக்கிடப் படுவதல்ல. இரண்டு புறக்கோடி மதிப்புக்களை மட்டுமே சார்ந்துள்ளது.

2. அது நிலையான அளவை அல்ல. கூறுகளுக்கிடையே அதிக வேறுபாட்டைக் காட்டும்.

3. திறந்த முனைப்பிரிவு இடைவெளிகள் உள்ள பரவலுக்கு வீச்சு காண முடியாது.

4. புறக்கோடி மதிப்புக்களால் பெரிதும் பாதிக்கப்படுகிறது. எடுத்துக்காட்டாக ஒரு பரீட்சையில் ஒரு மாணவன் 0 மதிப்பெண் பெற்றால் வீச்சு அதிக அளவில் வேறுபடலாம்.

முன்னர் வீச்சு $70-40 = 30$

தற்போது வீச்சு $70 - 0 = 70$

5. இது மிகவும் நம்பகமான அளவை அல்ல. ஏனெனில் அது வீச்சுக்கு இடைப்பட்ட எண்கள் பரவியுள்ள வகை நமக்குத் தெரிவிப்பதில்லை.

5. இது மிகவும் நம்பகமான அளவை அல்ல. ஏனெனில் அது வீச்சுக்கு இடைப்பட்ட எண்கள் பரவியுள்ள வகை நமக்குத் தெரிவிப்பதில்லை.

A	50	60	55	45	70	60	40	20
B	27	70	20	25	30	70	18	40
C	69	21	70	20	67	25	64	28

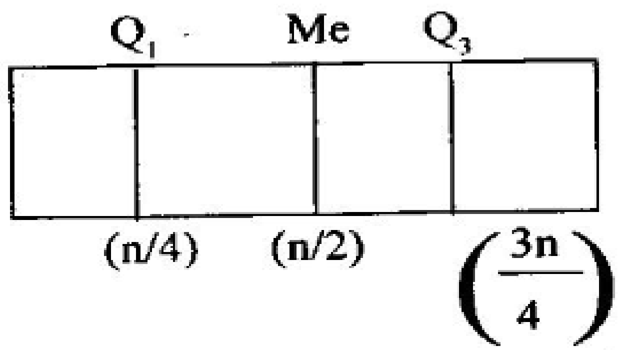
மூன்றிலும் வீச்சு 70 - 20 = 50 ஆகும். ஆனாலும் மதிப்புகள் பரவியுள்ள வகையில் மூன்றும் வேறுபடுவதைக் காணலாம்.

கால்மான விலக்கம்
(Quartile Deviation)

ஒரு பரவலின் அரைப் பகுதியை இருகால் பகுதிகளாகப் பிரிக்கும் எண்ணுக்கு முதற்கால்மானம் அல்லது கீழ்க்கால் மானம் எனப் பெயர். அவ்வாறே இரண்டாம் அரைப்பகுதியை இருகால் பகுதிகளாக பிரிக்கும் எண்ணை மூன்றாம் கால்மானம் அல்லது மேல் கால் மானம் என்கிறோம். மேல்கால்மானத்துக்கும் கீழ்கால்மானத்துக்கும் இடையில் உள்ள தூரம் இடைக்கால்மான வீச்சு (Inter quartile range) எனப்படும். இதுபாதிமான விலக்கம் அல்லது அரை இடைக்கால்மான வீச்சு (Semi - Inter quartile range) எனப்படும்.

$$\text{கால்மான விலக்கம் Q. D.} = \frac{Q_3 - Q_1}{2}$$

இதில் Q_3 = மூன்றாம் கால்மானம் ; Q_1 = முதல் கால்மானம். இடைநிலை, பரவலை இரு பகுதிகளாகவும்: கால்மானங்கள் நான்கு பகுதிகளாகவும் பிரிக்கும்.



இடைநிலையைக் காண்பதைப்போல Q_1 Q_3 ஆகியவற்றைக் கணக்கிட்டுவிடலாம். Q_1 Q_3 ஆகியவை ஓர் இயல்நிலைப்பரவலில் இரு பகுதியிலும் சம தூரத்தில் அமைந்து இருக்கும். (குறிப்பு: இரண்டாம் கால்மானமும் இடைநிலையும் ஒன்றுதான்.)

கால்மான விலக்கம் ஒரு முழு அளவை ஆகும். ஒப்பிட்டு நோக்கும் முகமாக சார்பு அளவை காண வேண்டும. அது கால்மான விலக்கக்கெழு (Coefficient of quartile deviation) எனப்படும்.

$$\boxed{\text{கால்மான விலக்கக்கெழு} = \frac{Q_3 - Q_1}{Q_3 + Q_1}}$$

மாதிரி: 1

பின்வரும் எண்களுக்கு கால்மான விலக்கம் காண்க.

10, 15, 13, 18, 8, 16, 17, 14, 20, 6, 22,

முதலில் ஏறுவரிசையில் வரிசைப்படுத்த வேண்டும்.

6, 8, 10, 13, 14, 15, 16, 17, 18, 20, 22

$Q_3 = 3\left(\frac{n+1}{4}\right)$ வது எண்

$Q_3 = 3\left(\frac{11+1}{4}\right)$

$= 9$ வது எண்.

$Q_3 = 18$

$Q_1 = \frac{n+1}{4}$ வது எண்.

$= \frac{11+1}{4} = 3$ வது எண்.

$Q_1 = 10$

$Q.\ D. = \frac{18-10}{2} = \frac{8}{2} = 4$

மாதிரி : 2

பின்வரும் பரவலில் கால்மான விலக்கமும் கால்மான விலக்கக் கெழுவும் காண்க.

6, 8, 10, 13, 14, 15, 16, 17, 18, 20, 22, 24

தீர்வு:

இங்கு விவரங்கள் வரிசைப்படுத்தப்பட்டு கொடுக்கப் பட்டிருக்கின்றன.

முதல் கால்மானம் $Q_1 = \left(\frac{n+1}{4}\right)$ வது எண்

$$n = 12 \,.\; Q_1 = \frac{12+1}{4} = \frac{13}{4} = 3\tfrac{1}{4} = 3.25$$

3- வது எண் = 10

∴ 3. 25- வது எண் = 3 வது எண் + (4-வது எண் - 3 வது எண்) x ¼

$$= 10 + (13 - 10) \times 1/4$$

$$= 10 + 3/4$$

$$Q_1 = 10.75.$$

மூன்றாம் கால்மானம் $Q_3 = 3 \left(\frac{n+1}{4}\right)$ வது எண்

$= 3\left(\frac{13}{4}\right) = 9.75$ வது எண்

9-வது எண் = 18; 10 -வது எண் = 20

$$\therefore Q_3 = 18 + (20 - 18) \times 3/4$$

$$= 18 + 2 \times 3/4 = 18 + 1.5$$

$$Q_3 = 19.5$$

கால்மான விலக்கம் $= \frac{Q_3 - Q_1}{2}$

$$= \frac{19.5 - 10.75}{2}$$

$$= \frac{875}{2} = 4.375$$

கால்மான விலக்கக்கெழு $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

$$= \frac{19.50 - 10.75}{19.50 + 10.75} = \frac{8.75}{30.4} = 0.28$$

மாதிரி : 3

பின்வரும் விவரங்களுக்கு கால்மான விலக்கக்கெழு காண்க.

மதிப்பெண்	30	40	50	60	70	80
மாணவர்கள் எண்ணிக்கை	2	6	10	8	3	2

X	f	cf
30	2	2
40	6	8
50	10	18
60	8	26
70	3	29
80	2	31

$Q_1 = \left(\frac{n+1}{4}\right)$ வது எண்

$= \frac{31+1}{4} = 8$ வது எண்

$Q_1 = 40$

$Q_3 = 3\left(\frac{n+1}{4}\right)$ வது எண் $3\left(\frac{31+1}{4}\right)$

$= 24$ வது எண்

$= 60$

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

$$= \frac{60 - 40}{2} = \frac{20}{2}$$

$$Q.D. = \frac{Q_3 - Q_1}{Q_3 + Q_1}$$

$$= \frac{60 - 40}{60 + 40} = \frac{20}{100} = 0.2$$

கணக்கு: 4

பின்வரும் அலைவெண் பரவலுக்கு கால்மான விலக்கம் காண்க.

CI	f	cf I
0-10	1	1
10-20	6	7
20-30	15	22
30-40	20	42
40-50	14	56
50-60	6	62
60-70	2	64

Q_1 = n/4 வது எண்

$= \frac{64}{4} = 16$

Q_1 உள்ளபிரிவு இடைவெளி 20-30

$$Q_1 = L + \frac{n/4 - cf}{15} \times c$$

$$= 20 + \frac{16-7}{15} \times 10$$

$$= 20 + \frac{90}{15} = 26$$

$$Q_3 = 3n/4 = 3\left(\frac{64}{4} = 48\right)$$

Q_3 பிரிவு இடைவெளி 40-50

$$Q_3 = L + \frac{3n/4 - cf}{f} \times i$$

$$= 40 + \frac{48-42}{14} \times 10$$

$$= 40 + \frac{60}{14} = 40 + 4.28$$

$$= 44.28$$

$$Q.D. = \frac{Q_3 - Q_1}{2}$$

$$= \frac{44.28 - 26}{2}$$

$$= \frac{18.28}{2} = 9.14$$

கால்மான விலக்கத்தின் நிறைகளும், குறைகளும்

நிறைகள்:

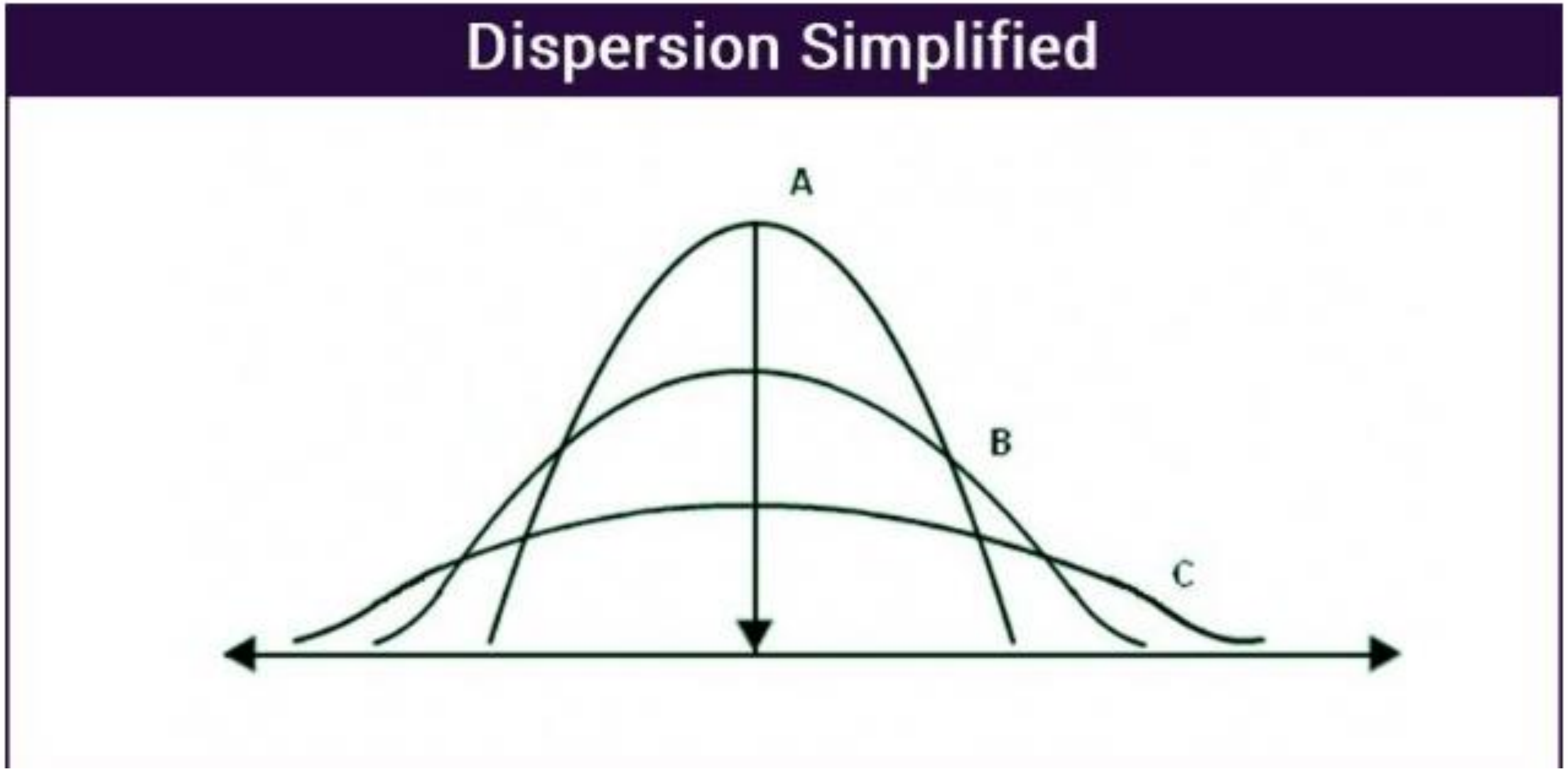
1. இது மிகவும் எளிதில் கணக்கிடக்கூடியது.
2. முதலிலும் ஈற்றிலும் வரும் எண்களால் பாதிக்கப்படாதது.
3. முதல் கடைசிப்பிரிவு இடைவெளிகள் தெரியாவிடினும் இதனைக் கணக்கிட்டுவிடலாம்.
4. இதனை வரைபடம் மூலமும் கண்டுபிடிக்கலாம்.
5. இது நடுவில் உள்ள பாதி அலைவெண்களது வீச்சைக் குறிக்கிறது. எனவே வீச்சைவிடச் சிறந்தது.
6. இது திட்டவட்டமான சூத்திரத்தால் வரையறுக்கப் பட்டுள்ளது.
7. புறக்கோடி மதிப்புக்கள் இதனைப்பாதிக்காது. எனவே மிகவும் கோட்டமுள்ள பரவலுக்கும் இதனைக் கணக்கிடலாம்.

குறைகள்:

1. இது எல்லா எண்களின் அடிப்படையிலும் கணக்கிடப்படுவதல்ல.
2. இது 50 சதவீதம் விவரங்களை மட்டுமே உள்ளடக்கி கணக்கிடப்படுகிறது. மேலுள்ள 25% எண்களும் தள்ளப்பட்டதால் இதனை பிரிவினை அளவை (Measure of Partition) என்கிறோம்.
3. கணக்கியல் முறைகளில் இது பயன்படாது.
4. இது நிலையான அளவை அல்ல. கூறுகளுக்கு கூறு அதிக வேறுபாட்டைக் காட்டும்.

Dispersion in Statistics

Dispersion is the state of getting dispersed or spread. Statistical dispersion means the extent to which a numerical data is likely to vary about an average value. In other words, dispersion helps to understand the distribution of the data.



Measures of Dispersion

In statistics, the measures of dispersion help to interpret the variability of data i.e. to know how much homogenous or heterogeneous the data is. In simple terms, it shows how squeezed or scattered the variable is.

Types of Measures of Dispersion

There are two main types of dispersion methods in statistics which are:

- Absolute Measure of Dispersion
- Relative Measure of Dispersion

Absolute Measure of Dispersion

An absolute measure of dispersion contains the same unit as the original data set. Absolute dispersion method expresses the variations in terms of the average of deviations of observations like standard or means deviations. It includes range, standard deviation, quartile deviation, etc.

Relative Measure of Dispersion

The relative measures of depression are used to compare the distribution of two or more data sets. This measure compares values without units. Common relative dispersion methods include:

1. Co-efficient of Range
2. Co-efficient of Variation
3. Co-efficient of Standard Deviation
4. Co-efficient of Quartile Deviation
5. Co-efficient of Mean Deviation

Co-efficient of Dispersion

The coefficients of dispersion are calculated (along with the measure of dispersion) when two series are compared, that differ widely in their averages. The dispersion coefficient is also used when two series with different measurement unit, are compared. It is denoted as C.D.

The common coefficients of dispersion are:

C.D. In Terms of	**Coefficient of dispersion**
Range	C.D. = $(X_{max} - X_{min}) / (X_{max} + X_{min})$
Quartile Deviation	C.D. = $(Q3 - Q1) / (Q3 + Q1)$
Standard Deviation (S.D.)	C.D. = S.D. / Mean
Mean Deviation	C.D. = Mean deviation/Average

The types of absolute measures of dispersion are:

1. **Range:** It is simply the difference between the maximum value and the minimum value given in a data set. Example: 1, 3,5, 6, 7 => Range = 7 -1= 6
2. **Variance:** Deduct the mean from each data in the set then squaring each of them and adding each square and finally dividing them by the total no of values in the data set is the variance. Variance $(\sigma^2)=\sum(X-\mu)^2/N$
3. **Standard Deviation:** The square root of the variance is known as the standard deviation i.e. S.D. = $\sqrt{\sigma}$.
4. **Quartiles and Quartile Deviation:** The quartiles are values that divide a list of numbers into quarters. The quartile deviation is half of the distance between the third and the first quartile.
5. **Mean and Mean Deviation:** The average of numbers is known as the mean and the arithmetic mean of the absolute deviations of the observations from a measure of central tendency is known as the mean deviation (also called mean absolute deviation).

- The **measure of dispersion** shows the scatterings of the data.
- It tells the variation of the data from one another and gives a clear idea about the distribution of the data.
- The measure of dispersion shows the homogeneity or the heterogeneity of the distribution of the observations.
- **Characteristics of Measures of Dispersion**
- A measure of dispersion should be rigidly defined
- It must be easy to calculate and understand
- Not affected much by the fluctuations of observations
- Based on all observations

- **Classification of Measures of Dispersion**
- The measure of dispersion is categorized as:
- (i) An absolute measure of dispersion:
- The measures which express the scattering of observation in terms of distances i.e., range, quartile deviation.
- The measure which expresses the variations in terms of the average of deviations of observations like mean deviation and standard deviation.
- (ii) A relative measure of dispersion:
- We use a relative measure of dispersion for comparing distributions of two or more data set and for unit free comparison. They are the coefficient of range, the coefficient of mean deviation, the coefficient of quartile deviation, the coefficient of variation, and the coefficient of standard deviation.

- **Range**
- A range is the most common and easily understandable measure of dispersion. It is the difference between two extreme observations of the data set. If X max and X min are the two extreme observations then
- Range = X max – X min
- **Merits of Range**
- It is the simplest of the measure of dispersion
- Easy to calculate
- Easy to understand
- Independent of change of origin
- **Demerits of Range**
- It is based on two extreme observations. Hence, get affected by fluctuations
- A range is not a reliable measure of dispersion
- Dependent on change of scale

- **Quartile Deviation**
- The quartiles divide a data set into quarters. The first quartile, (Q1) is the middle number between the smallest number and the median of the data. The second quartile, (Q2) is the median of the data set. The third quartile, (Q3) is the middle number between the median and the largest number.
- Quartile deviation or semi-inter-quartile deviation is
- $Q = \frac{1}{2} \times (Q3 - Q1)$
- **Merits of Quartile Deviation**
- All the drawbacks of Range are overcome by quartile deviation
- It uses half of the data
- Independent of change of origin
- The best measure of dispersion for open-end classification
- **Demerits of Quartile Deviation**
- It ignores 50% of the data
- Dependent on change of scale
- Not a reliable measure of dispersion