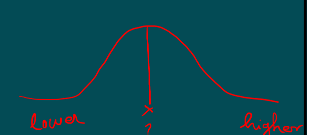
Measure of Central Tendency (Part 1)

Basics Concepts
Types
Dispersion, skewness, kurtosis

Biostatistics & Research Methodology
B Pharm 8th Sem | M. Pharm. | PhD

Measures of Central Tendency

- **Measures of central tendency**- Statistical Average
- It tell us the point about which items have a tendency to collection. Such a measure is considered as the most representative figure/Value for the entire mass of data.
- Represent the single value to describe the whole dataset
- This representative value is called the measure of central tendency
- Mean, Median and Mode, are most common Statistical Average



Measures of Central Tendency

Classification

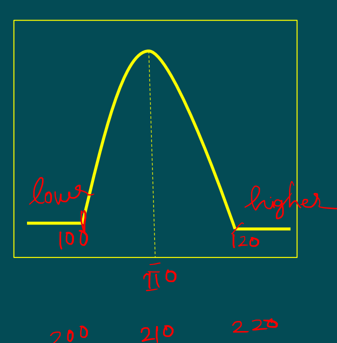
- Average
  - Mathematical Averages
    - Arithmetic Mean
      - Simple Means
      - Weighted Mean
    - Geometric Mean
    - Harmonic Mean
  - Average of Position
    - Median
    - Mode
    - Percentile
    - Deciles
    - Quartiles

Measures of Central Tendency

Objectives

- To get the single value that describe the whole data group (i.e., average blood sugar level of whole group)
- To facilitate compare between two different group.

SN	Group A	Group B
1 x_1	100	210
2 x_2	120	200
3 x_3	110	220
Mean	**110** Vs	**210**

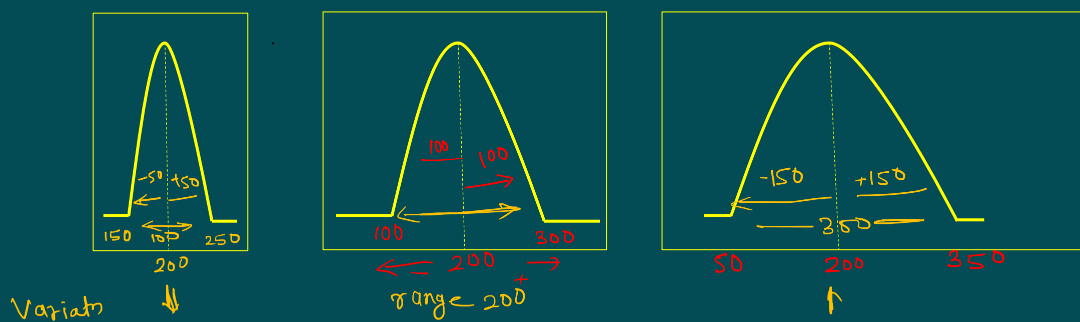


- It offer the base for computing other measure like variation, dispersion, skewness, kurtosis, etc.

Measures of Central Tendency

Objectives

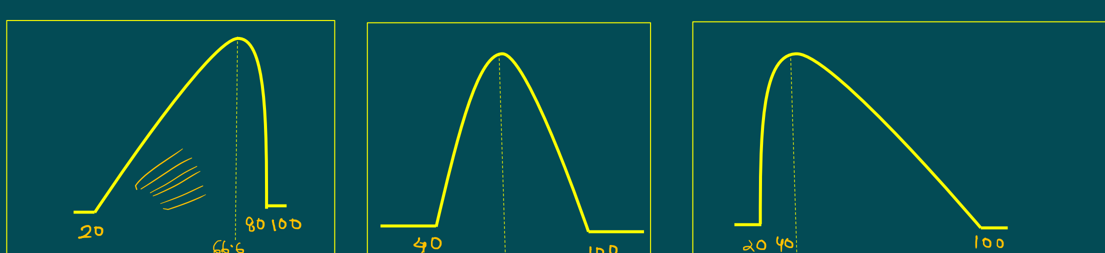
- It offer the base for computing other measure like variation, **dispersion**, skewness, kurtosis, etc.



Measures of Central Tendency

Objectives

- It offer the base for computing other measure like variation, dispersion, **skewness**, kurtosis, etc.



Left Skewed (-ve) Right Skewed (-ve)

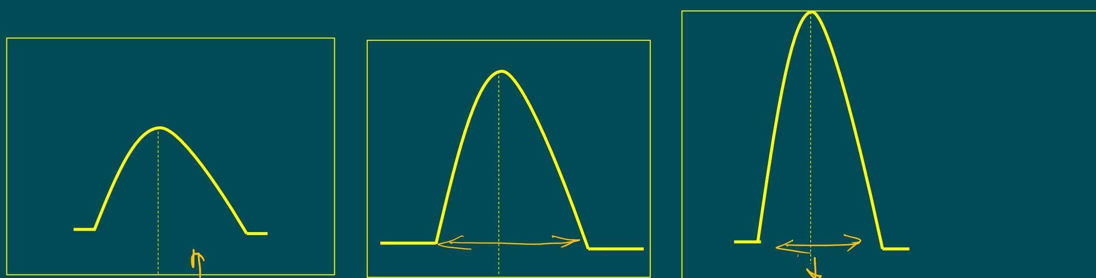
m = 200/3 = 66.6

180/3 = 60

m = 160/3 = 53.—

Measures of Central Tendency

- **Objectives**
- It offer the base for computing other measure like variation, dispersion, skewness, **kurtosis**, etc.

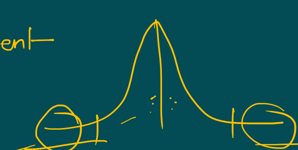


Platykurtic MesoKurtic Leptokurtic

Measures of Central Tendency

- Easy to Understand
- Simple to computation and Comparison
- Based on all dataset
- Not affected by extreme observations
- Sampling stability (random sampling)

N = 100 Student



Measure of Central Tendency
(Part 2)

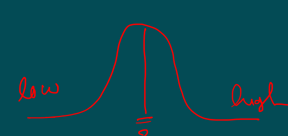
- ✓ Mean
- ✓ Median
- ✓ Mode

Biostatistics & Research Methodology

B Pharm 8th Sem | M. Pharm. | PhD

Measures of Central Tendency

- **Measures of central tendency**- Statistical Average
- Represent the single value to describe the whole dataset
- This representative value is called the measure of central tendency
- Mean – Arithmetic mean- Average of deta
- Median – Mid point of the data
- Mode- Most frequent data



$\frac{(n+1)}{2}^{th} = \frac{6}{2} = 3^{rd}$

1, 3, 8, 10, 12

Example: 1, 3, 8, 8, 10, 12

$x_1 \; x_2 \; x_3 \; x_4 \; x_5 \; x_6$ $\quad N = 6$

- **Mean- 42/6 = 7** $\quad \bar{x} = \frac{\Sigma x}{N} = \frac{1+3+8+8+10+12}{6} =$
- **Median- 8**
- **Mode- 8**

Mean

Mean is average of the given numbers which is calculated by sum of all the values of data divided by the total number of values.

$$\text{Mean } \bar{X} = \frac{\sum Xi}{n} = \frac{X1+X2+X3\ldots\ldots+Xn}{n}$$

$\bar{X}$ = The symbol we use for mean (pronounced as *X* bar)
$\sum$ = Symbol for summation
Xi = Value of the *i*th item *X*, *i* = 1, 2, …, *n*
n = total number of items

SN	Group A	Group B
1 X_1	100	210
2 X_2	120	200
3 X_3	110	220
Mean	**110**	**210**

Mean (100+120+110)/3 (210+200+220)/3
330/3 = 110 630/3 = 210

Mean

Formula for individual data

$$\text{Mean } \bar{X} = \frac{\sum Xi}{n} = \frac{X1+X2+X3\ldots\ldots+Xn}{n} \quad \text{Or} \quad \bar{X} = A + \frac{\sum d}{n}$$, where A= Assumed mean d = X-A

Example 1: calculate the mean production of tablet by a tab punching machine, production of tablet/day in thousand- 120, 100, 110, 150, 120, 100, 120

1 2 3 4 5 6 7

N=7

mean= (120 + 100 + 110 + 150 +120 +100 +120)/ 7

820/7 = 117.15 thousand per day

Mean

Formula for individual data

Example 1: calculate the mean production of tablet by a tab punching machine, production of tablet/day in thousand- 120, 100, 110, 150, 120, 100, 120,

SN	Per Day Production (in Thousand) (X)	d (X-120)
1	120 − 120	0
2	100 −120	-20
3	110	-10
4	150	30
5	120	0
6	100	-20
7	120	0
Sum		-20

$\bar{X} = A + \frac{\sum d}{n}$, where A= Assumed mean d = X-A

$\bar{X} = A + \frac{\sum d}{n}$

= 120 + (-20/7)

= 120-2.85

= 117.15

= 117.15 Thousand per day

Mean

Formula for Discrete data

$\bar{X} = \frac{\sum_{i=1}^{n} fiXi}{\sum_{i=1}^{n} fi}$ Or $\bar{X} = A + \frac{\sum fd}{N}$, where A= Assumed mean d = X-A

$\sum_{i=1}^{n} fi = \mathrm{N}$ Sum of all frequency

Example 2. Hypertensive patient per family in a village, calculate the arithmetic mean (avg patient per family)

SN	No. of HTN Patients (X)	No. of Family (F)	FX
1	0	50	0
2	1	20	20
3	2	70	140
4	3	10	30
	SUM	150 Σfx	190

$X = \frac{\sum_{i=1}^{n} fiXi}{\sum_{i=1}^{n} fi} = \frac{\sum fx}{N}$

=190/150

=1.26

Mean

Formula for Continuous data

$$\bar{X} = \frac{\sum_{i=1}^{n} fiMi}{\sum_{i=1}^{n} fi} \quad \text{Or} \ \frac{\Sigma Fm}{N} \qquad \bar{X} = A + \frac{\sum fd}{N}$$, where A= Assumed mean d = M-A

m = mid point class interval

$\sum_{i=1}^{n} fi$ = N Sum of all frequency

Example 3. Hypertensive patient per age group in a village, calculate the arithmetic mean (avg age of patient)

	Age (Y)	No. of Patient (F)	M	FM
1	0-20	0	10	0
2	20-40	10	30	300
3	40-60	100	50	5000
4	60-80	80	70	5600
	SUM	ΣF or N 190	ΣFm	10900

$$X = \frac{\sum_{i=1}^{n} fiMi}{\sum_{i=1}^{n} fi}$$

=10900/190

=57.36

Mean

Formula for Continuous data

Example 4. Calculate the avg accident per week in the given data

	No of Accidents	No. of Week (F)	M	FM
1	0-10	15	5	75
2	10-20	10	15	150
3	20-30	20	25	500
4	30-40	7	35	245
	SUM	N ΣF 52		ΣFm 970

$$X = \frac{\sum_{i=1}^{n} fiMi}{\sum_{i=1}^{n} fi}$$

=970/52

=18.65

Mean

Example from Question paper

5. Find the mean of a sample of reported cases of mumps in school children

Blood LDL (X)	52	58	60	65	68	70	75
No. of Patients (F)	7	5	4	6	3	3	2

ΣF = 30
N

FX = 364 290 240 390 204 210 150 = 1848
ΣFX

$\bar{X} = \frac{\sum fx}{N}$

= 1848/30

= 61.6

								SUM
Blood LDL (X)	52	58	60	65	68	70	75	
No. Pat. (F)	7	5	4	6	3	3	2	30
D (X-60)	-8	-2	0	5	8	10	15	
FD	-56	-10	0	30	24	30	30	48

ΣF or N

ΣFd

d = X - A

$\bar{X} = A + \frac{\sum fd}{N}$

= 60 + (48/30)

= 60+1.6

= 61.6

Median

- Median represents the mid-value of the given set of data when arranged in a particular order
- Given that the data collection is arranged in ascending or descending order, the following method is applied:
- If number of values or observations in the given data is odd, then the median is given by **[(n+1)/2]th** observation.

  $\frac{5+1}{2} = \frac{6}{2} = 3^{rd}$

  8, 2, 5, 6, 1 → 1 2 5 6 8
  median

- If in the given data set, the number of values or observations is even, then the median is given by the average of (n/2)th and [(n/2)+1]th observation. = 5.5 median

  $\frac{\left(\frac{n}{2}\right) + \left(\frac{n}{2}+1\right)}{2} = \frac{3^{rd} + 4^{th}}{2}$

  $\frac{5+6}{2} = 5.5$

  1, 4, 5, 6, 7, 8

  avg = 5.5

  $\frac{12}{2} = 6$

Mode

- The most frequent number occurring in the data set is known as the mode.

1, 2, 2, 3, 4, 5, 6 = mode = 2
1, 2, 2, 3, 4, 4, 6 = mode = 2, 4
1, 2, 2, 2, 3, 3, 3, 4, 5, 5, 5, 6 mode = 2, 3, & 5

discrete data

SN	No. of HTN Patients (X)	No. of Family (F)
1	0	50
2	1	20
3	2	70
4	3	10
	SUM	150

mode = 2

Mode

- For Grouped Data.

Marks Obtained	No of Student
10-20	5 f_0
20-30 —10	12 = modal class f_1
30-40	8 f_2
40-50	5

$$Mode = l + \left(\frac{f_1 - f_0}{2f_1 - f_0 - f_2}\right) \times h$$

Where,
l = lower limit of the modal class — having high F
h = size of the class interval
f_1 = frequency of the modal class
f_0 = frequency of the class preceding the modal class
f_2 = frequency of the class succeeding the modal class

24-5-8

Mode = 20 + [(12-5) / (2x12 – 5 – 8)] x 10
20 + [7/11] x 10
20 + (0.63 x 10)
20 + 6.3
26.3

Measures of Dispersion
(Part 1)

BASIC CONCEPTS OF DISPERSION

Biostatistics & Research Methodology
B Pharm 8th Sem | M. Pharm. | PhD

Measures of Dispersion

- Dispersion means spread or distribution of data
- Statistical dispersion means variation form average value
- Dispersion is important for comparing the dataset/group

A

B

Tab production rate

$\bar{x} = 100K$

1 - 100K
2 - 200K H
3 - 50K
4 - 50K - L

$m = \frac{400K}{4} = 100K$

rang: 150

$\bar{x} = 100K$

1 100
2 0 — L
3 200 H
4 - 100

m = 400/4 = 100K

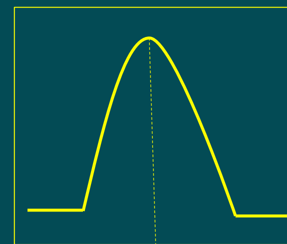
rang: 200 r

Measures of Dispersion

- Dispersion means spread or distribution of data
- Statistical dispersion means variation form average value
- Dispersion is important for comparing the dataset/group

A

B



5 Student 10 Lakhs

1 – 10
2 12
3 8
4 –10
5 –10

variation

5 Student 10 Lakhs

1 – 5
2 – 15
3 – 25
4 – 5
5 – 10

Measures of Dispersion

Significance of the Dispersion

	Lab 1 Tablet Machine A	Lab 2 Tablet Machine B	Lab 3 Tablet Machine C
d_1	300	200	50
d_2	300	200	150
d_3	500	600	800
d_4	500	600	600
$\bar{X}$	**400**	**400**	**400**

range: 500 − 300 = 200

100 − | + 100
400

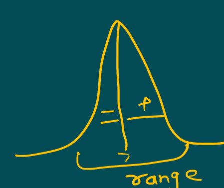
mean ± SD = 400 ± 100

range 600 − 200 = 400

200 − | + 200

= 400 ± 200

range: 800 − 50



Measures of Dispersion

Significance of the Dispersion

- Dispersion indicates the distribution of data
- It determine the reliability of an average
- It helps to control the variability
- It helps to compare the multiple group in respect to variability
- Also useful for other statistical measure

Tablet Machine A	Tablet Machine B	Tablet Machine C
300	200	50
300	200	150
500	600	800
500	600	600
400	**400**	**400**

good poor v poor

Measures of Dispersion

Measures of Dispersion

Absolute Measures of Dispersion
- Range
- Variance
- Standard Deviation
- Mean Deviation
- Quartile Deviation
- Lorenz Curve

Univ. 1 Studen = $\frac{40}{50}$ = 80 % Abs. Relative data

Relative measure of Dispersion
- Co-efficient of Range
- Co-efficient of Variance
- Co-efficient of Standard Deviation
- Co-efficient of Mean Deviation
- Co-efficient of Quartile Deviation

Univ. 2 Stud 2 = $\frac{50}{100}$ = 50 %

Measures of Dispersion

Property of Measures of Dispersion

- Simple and easy to understand
- Easy to compute and compare
- Rigidity defined
- Based on all date and not affected by extreme observation
- Sample stability

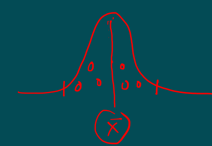
Measure of Dispersion
(Part 2)

BASICS OF RANGE

Biostatistics & Research Methodology
B Pharm 8th Sem | M. Pharm. | PhD

Measures of Dispersion

Measures of Dispersion		
	Absolute Measures of Dispersion	Range
		Variance
		Standard Deviation
		Mean Deviation
		Quartile Deviation
		Lorenz Curve
	Relative measure of Dispersion	Co-efficient of Range
		Co-efficient of Variance
		Co-efficient of Standard Deviation
		Co-efficient of Mean Deviation
		Co-efficient of Quartile Deviation



RANGE

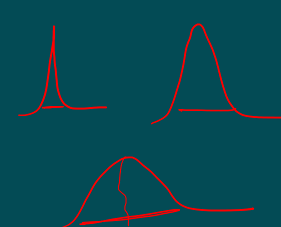
- Simplest method for determining measures of Dispersion
- Difference between smallest and Largest Value given in dataset

Range = (Largest-Smallest)

Coefficient of Range = (L-S)/(L+S)

- **Individual Data**

Day	Tablet Machine A	Tablet Machine B
1	300	200
2	300	200
3	500	600
4	500	600
Mean	**400**	**400**



A

Range = L-S
= 500-300 = **200**

Coefficient of Range = L-S/L+S
= 200/800 = **0.25**

B

Range = L-S
= 600-200 = **400**

Coefficient of Range = L-S/L+S
= 400/800 = **0.5**

RANGE

Discrete Data

SN	No. of HTN Patients (X)	No. of Family (F)	FX
1	0 — S	50	0
2	1	20	20
3	2	70	140
4	3 — L	10	30
	SUM	Σf = N 150	Σfx 190

$$\bar{X} = \frac{\sum fX}{N}$$

=190/150

mean =1.26

Range = L-S

= 3-0 = **3**

Coefficient of Range = L-S/L+S

= 3/3 = **1**

								Sum
Marks	S 5	10	11	15 L	8	7	9	
No. Student. (F)	10	10	5	1	14	10	10	60
FX	50	100	55	15	112	70	90	492

mean =492/60 = 8.2 Range = L-S= 15-5 = 10 Coefficient of Range = L-S/L+S

= 15/20= 0.75

RANGE

Contineous Data

	Age (Y)	No. of Patient (F)	M	FM
2	20-40 S	10	30 S	300
3	40-60	100	50	5000
4	60-80 L	80	70 L	5600
	SUM	190		10900

$$\bar{X} = \frac{\sum fM}{N}$$

=10900/190

mean =57.36

Range = L-S

= 70-30 = **40**

A

Coefficient of Range = L-S/L+S

= 40/100

= **0.4**

B

Range = L-S

= 80-20 = **60**

Coefficient of Range = L-S/L+S

= 60/100

= **0.6**

RANGE

MERITS

- Easy and Simple
- Rapid Calculation
- Very Quick picture to variability

1 2 4 800 900 1000 2000

S L

2000 − 1 = 1999

DEMERITS

- Not include every data
- Less Accuracy
- It not tell any thing about character of the distribution
- Can't be computed in case of open end distribution

S − 10 − 20

20 − 30

30 − 40

40 − 50

> 50

Measures of Dispersion (Part 3)

Standard Deviation and Variance

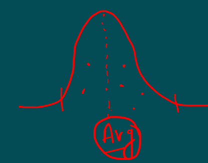
Biostatistics & Research Methodology

B Pharm 8th Sem | M. Pharm. | PhD

Measures of Dispersion

Measures of Dispersion

- Absolute Measures of Dispersion
  - Range
  - Variance
  - Standard Deviation
  - Mean Deviation
  - Quartile Deviation
  - Lorenz Curve
- Relative measure of Dispersion
  - Co-efficient of Range
  - Co-efficient of Variance
  - Co-efficient of Standard Deviation
  - Co-efficient of Mean Deviation
  - Co-efficient of Quartile Deviation



STANDARD DEVIATION

- Most commonly used to determine the dispersion
- Measures of Absolute dispersion
- SD directly proportional to dispersion (Greater SD = Greater Dispersion)
- SD define as the square root of variance, which denote as sigma (σ)
  - $\sigma = \sqrt{\text{Variance}}$
  - $\sigma^2 = \text{Variance}$
- Variance: The average of the **squared** differences from the Mean.

STANDARD DEVIATION

- Example (Individual data): calculate the mean, Variance and SD
- Marks of students (N = 5): 8, 7, 8, 7, 10
  - Mean = (8+7+8+7+10)/5 = 40/5= 8
  - Variance (σ^2) = $\frac{\sum(X1-mean)^2+(X2-mean)^2+\cdots\ldots+(Xn-mean)^2}{N}$

    = (0+1+0+1+4)/5 = 6/5 = 1.2

  SD (σ) = √Variance

  = √1.2 = 1.09

S N	Marks	Diference from mean (X-m)	$(X\text{-}m)^2$
1	8	0	0
2	7	-1	1
3	8	0	0
4	7	-1	1
5	10	2	4
m	**8**		Sum= 6

***If we used population then we divided by N in calculation of variance**

***If we used a sample the we devided by N-1 in calculation of variance**

For Sample:

Variance- 6/4 = 1.5 **SD = √ 1.5 = 1.22**

STANDARD DEVIATION

- **For Population, SD (σ) =** $\sqrt{\sum_1^n (Xi - mean)^2 / N}$
- **For Sample, SD (σ) =** $\sqrt{\sum_1^n (Xi - mean)^2 / N\text{-}1}$

For Individual data: $SD = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$

For Discrete data: $SD = \sqrt{\frac{\Sigma Fd^2}{N} - \left(\frac{\Sigma Fd}{N}\right)^2}$

For Contineous data $SD = \sqrt{\frac{\Sigma Fd^2}{N} - \left(\frac{\Sigma Fd}{N}\right)^2} \times i$

Coefficient of Variation (CV or % CV)

CV = (σ/mean) x 100

STANDARD DEVIATION

- *For Individual data:* $SD = \sqrt{\frac{\Sigma d^2}{N} - \left(\frac{\Sigma d}{N}\right)^2}$
- **Example:** Calculate the SD of given data- 2, 4, 8, 10, 12, 16

X	d (X-A)	d2
2	-8	64
4	-6	36
8	-2	4
10	0	0
12	2	4
16	6	36
M = 8.6	-8	144

$$SD = \sqrt{\frac{144}{6} - \left(\frac{-8}{6}\right)^2}$$

$$SD = \sqrt{24 - (-1.33)^2}$$

$$SD = \sqrt{24 - 1.76}$$

$$SD = \sqrt{22.24}$$

$$SD = 4.71$$

CV = (4.71/8.66) x 100 = 0.543 x 100 = 54.3

STANDARD DEVIATION

- *For Discrete data:* $SD = \sqrt{\frac{\Sigma Fd^2}{N} - \left(\frac{\Sigma Fd}{N}\right)^2}$
- **Example:**

Marks	No Students (F)	FX	d (X-A)	d^2	Fd	Fd^2
4	12	48	-2	4	-24	48
5	10	50	-1	1	-10	10
6	8	48	0	0	0	0
8	10	80	2	4	20	40
10	10	100	4	16	40	160
	50	**326**			**26**	**258**

$$SD = \sqrt{\frac{258}{50} - \left(\frac{26}{50}\right)^2}$$

$$SD = \sqrt{5.16 - (0.52)^2}$$

$$SD = \sqrt{5.16 - 0.27}$$

$$SD = \sqrt{4.89}$$

$$SD = 2.21$$

CV = (2.21/6.52) x 100 = 0.338 x 100 = 33.8

STANDARD DEVIATION

- **For Continuous data** $SD = \sqrt{\frac{\Sigma Fd^2}{N} - \left(\frac{\Sigma Fd}{N}\right)^2}$ **x i**
- **Example:** Calculate the SD of given data

Marks	No Students (F)	M	FM	d (M-25)/10	d^2	Fd	Fd^2
0-10	12	5	60	-2	4	-24	48
10-20	10	15	150	-1	1	-10	10
20-30	8	25	200	0	0	0	0
30-40	10	35	350	1	1	10	10
40-50	10	45	450	2	4	20	40
	50		1210			-4	108

$SD = \sqrt{\frac{108}{50} - \left(\frac{-4}{50}\right)^2}$ x 10

$SD = \sqrt{2.16 - (-0.08)^2}$ x 10

$SD = \sqrt{2.16 - 0.0064}$ x 10

$SD = \sqrt{2.16}$ x 10

$SD = 1.46\ X\ 10 = 14.6$

CV = (14.6/24.2) x 100 = 0.603 x 100 = 60.3

Measure of Dispersion
(Part 4)

Mean Deviation & Coefficient of Mean Deviation

Biostatistics & Research Methodology

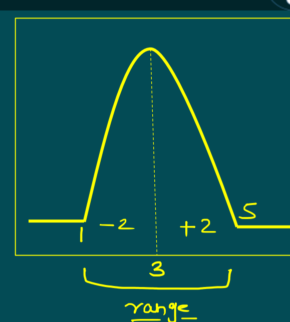
B Pharm 8th Sem | M. Pharm. | PhD

Measures of Dispersion

Measures of Dispersion		
	Absolute Measures of Dispersion	Range
		Variance
		Standard Deviation
		Mean Deviation
		Quartile Deviation
		Lorenz Curve
	Relative measure of Dispersion	Co-efficient of Range
		Co-efficient of Variance
		Co-efficient of Standard Deviation
		Co-efficient of Mean Deviation
		Co-efficient of Quartile Deviation

MEAN DEVIATION

- Mean Deviation- Average deviation from the mean value in given dataset.
- Average difference between the item in a distribution vs mean/median



Mean Deviation = Σ (| Xi – Avg |)/N

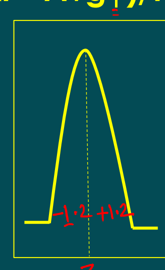
mean/ med

N = 5

= 6/5

= 1.2

X	Difference from mean $|X-A|$
1 − 3 = |−2|	2
2	1
3	0
4	1
5	2
Mean = 3 = 15/5	6 = $\Sigma\|X-A\|$



MEAN DEVIATION

- Individual Data $MD = \frac{\sum |D|}{N}$ D = X−A Avg mean/median
- Discrete Data $MD = \frac{\sum f|D|}{N}$
- Continuous Date $MD = \frac{\sum f|D|}{N}$
- Coefficient of MD = MD/median

MEAN DEVIATION

- Individual Data

X	|D| = |X-A|
1	2
2.	1
3. — median	0
4	1
5	2
Sum = 15	Sum= 6
A(Median) = 3	

$MD = \frac{\sum |D|}{N}$

|D| = |X-A|

A = Statistical Average (Mean or Median)

$\left(\frac{n+1}{2}\right)^{th}$

$\frac{5+1}{2} = 3^{rd}$

MD = 6/5 = 1.2

Coefficient of MD = MD/Median

= 1.2/3 = 0.4

MEAN DEVIATION

- Discrete Data

Marks X	No Students (F)	CF	|D| = |X-A|	F|D|
4	12	12	2	24
5	10	22	1	10
6	8	30	0	0
8	10	40	2	20
10	10	50	4	40
	Sum = 50 N			94
A = 6				

$$MD = \frac{\sum f|D|}{N}$$

|D| = |X-A|

A = Statistical Average (Mean or Median)

Median = N+1/2 = 50+1/2 = 25.5th $\frac{25^{th} + 26^{th}}{2}$

Median = 6+6/2 = 12/2 = 6

MD = 94/50 = 1.88

Coefficient of MD = 1.88/6 = 0.31

MEAN DEVIATION

- Continuous Data

$$MD = \frac{\sum f|D|}{N}$$

mid value of class interval

|D| = |M-A|

A = Statistical Average (Mean or Median)

Median = L + [(N/2) – CF]/ f x i

N/2 = 50/2 = 25 CF = 22, F = 8, i = 10, L = 20

Median = $20 + \frac{25-22}{8} \times 10$

Median = $20 + 0.375 \times 10$ = 20+3.75 = 23.75

MD = 647.5/50 = 12.95

Coefficient of MD = 12.95/23.75 = 0.54

Marks m	No Students (F)	CF	M	|D| = |M-A|	F|D|
0-10	12	12	5	18.75	225
10-20	10	22	15	8.75	87.5
20-30 L	8 F	30	25	1.25	10
30-40	10	40	35	11.25	112.5
40-50	10	50	45	21.25	212.5
	50			Σf|D|	647.5
A = 23.75					

Thanks for Watching

Subscribe my YouTube Channel