Unit 5: Statistics

Applied Mathematics — 4320001 / DI02000011

Milav Dabgar

Lecturer, Dept of ECE
GP Palanpur

April 2, 2026

Unit 5 — Table of Contents

Syllabus Coverage

Weightage: ≈12 marks (17%) | Hours: 9–10

5.1 Introduction & Key Definitions

Why Statistics?

Key Measures in This Unit

Mean ($\bar{x}$)	Central tendency (average)
Mean Deviation (MD)	Average distance from mean
Standard Deviation (σ)	Spread / dispersion
Variance (σ^2)	Square of SD
Coeff. of Variation (CV)	Relative spread (%)
Range	Max − Min

Notation

n = number of observations f_i = frequency x_i = value / midpoint
$N = \sum f_i$ (total frequency) $\bar{x}$ = mean

MCQ Bank — 1-Mark Questions

All MCQs at a Glance

Question	Answer	Paper
Mean of first 5 natural numbers	3	W'22, W'23
Mean of $11, x, 19, 21, y, 29 = 20$; $x + y = ?$	40	W'22
Mean of $39, 23, 58, 47, 50, 16, 61$	47	W'23
Mean of first 10 natural numbers	5.5	Su'24
Range of $17, 15, 25, 34, 32$	19	Su'24
Mean of $1, 3, 5, 7, 9$	5	W'24
Mean of $15, 7, 6, a, 3 = 4$; $a = ?$	-11	W'24
Mean of first 5 even natural numbers	6	Su'25

5.2 Mean: Direct Method

Mean Formulas

Ungrouped Data

$$\bar{x} = \frac{\sum x_i}{n} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Grouped (Frequency Distribution)

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{N}$$

where x_i = **midpoint** of class $= \dfrac{\text{lower} + \text{upper}}{2}$

Combined Mean

$$\bar{x}_{12} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2}$$

Mean Example 1 — Ungrouped (GTU MCQ practice)

Q: Mean of $1, 3, 5, 7, 9$

$$\bar{x} = \frac{1+3+5+7+9}{5} = \frac{25}{5} = \boxed{5}$$ (W'24)

Q: Mean of first 5 natural numbers

$$1, 2, 3, 4, 5 \Rightarrow \bar{x} = \frac{15}{5} = \boxed{3}$$ (W'22, W'23)

Q: If mean of $15, 7, 6, a, 3$ is 4, find a

$$\frac{15+7+6+a+3}{5} = 4 \Rightarrow 31 + a = 20 \Rightarrow \boxed{a = -11}$$ (W'24)

Mean Example 2 — Frequency Table (GTU W'22)

Q: Find mean x_i: 92,93,97,98,102,104 f_i: 3,2,3,2,6,4

x_i	f_i	$f_i x_i$
92	3	276
93	2	186
97	3	291
98	2	196
102	6	612
104	4	416
Total	**20**	**1977**

$$\bar{x} = \frac{1977}{20} = \boxed{98.85}$$

Mean Example 3 — Grouped (GTU Su'24)

Q: Find mean (Age 20–59, staff data)

Class	f_i	mid x_i	$f_i x_i$
20–24	5	22	110
25–29	7	27	189
30–34	9	32	288
35–39	11	37	407
40–44	10	42	420
45–49	8	47	376
50–54	6	52	312
55–59	4	57	228
Total	**60**		**2330**

$$\bar{x} = \frac{2330}{60} = \boxed{38.83}$$

Combined Mean (GTU Su'24)

Q: 25 obs. with mean 50 and 75 obs. with mean 60 — find combined mean

$$\bar{x}_{12} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2} = \frac{25 \times 50 + 75 \times 60}{25 + 75}$$
$$= \frac{1250 + 4500}{100} = \frac{5750}{100} = \boxed{57.5}$$

Q: If mean of $15, 7, 6, a, 3$ is 7, find a (W'22)

$$\frac{31 + a}{5} = 7 \Rightarrow 31 + a = 35 \Rightarrow \boxed{a = 4}$$

5.3 Mean: Assumed Mean & Step Deviation

Shortcut methods for grouped frequency data

Assumed Mean & Step Deviation — Formulas

Assumed Mean Method

Choose assumed mean A (usually central value).

$$d_i = x_i - A \quad \text{then} \quad \bar{x} = A + \frac{\sum f_i d_i}{N}$$

Step Deviation Method

Choose class width h and let $u_i = \dfrac{x_i - A}{h}$.

$$\bar{x} = A + h \cdot \frac{\sum f_i u_i}{N}$$

When to use

Use step deviation when all class widths are equal — it gives smaller numbers to work with.

Step Deviation Example — GTU W'22 (4-mark)

Find mean marks: $A = 38,\ h = 5$

Class	f	x	$u = \frac{x-38}{5}$	fu
21 – 25	8	23	-3	-24
26 – 30	10	28	-2	-20
31 – 35	24	33	-1	-24
36 – 40	30	38	0	0
41 – 45	12	43	1	12
46 – 50	16	48	2	32
Total	**100**			$\mathbf{-24}$

$$\bar{x} = 38 + 5 \times \frac{-24}{100} =$$

$38 - 1.2 = \boxed{36.8}$

Assumed Mean Example — GTU W'23

Find mean score; class 0–120, $A = 60$, $h = 20$

Class	f	x	$u = \frac{x-60}{20}$	fu
0 – 20	26	10	-2.5	-65
20 – 40	31	30	-1.5	-46.5
40 – 60	35	50	-0.5	-17.5
60 – 80	42	70	0.5	21
80 – 100	82	90	1.5	123
100–120	71	110	2.5	177.5
Total	**287**			**192.5**

$$\bar{x} = 60 + 20 \times \frac{192.5}{287} = 60 + 13.43 = \boxed{73.43}$$

Missing Frequency — GTU Su'23

Mean$= 19$; f_i: 2,4,7,f,8,4,3 for x_i: 6,10,14,18,24,28,30 (3 marks)

$N = 28 + f \quad \sum f_i x_i = 12 + 40 + 98 + 18f + 192 + 112 + 90 = 544 + 18f$

$$\bar{x} = 19 \Rightarrow \frac{544 + 18f}{28 + f} = 19$$

$544 + 18f = 532 + 19f \Rightarrow \boxed{f = 12}$

Missing frequency (W'24): mean$= 100$

Use same method: set up $\bar{x} = \dfrac{\sum fx}{N} = 100$ and solve for unknown x.

Mean from Freq Table — GTU Su'25

x_i: 17,20,23,26,29,32,35; f_i: 6,8,15,17,10,7,4 (3 marks)

$N = 6 + 8 + 15 + 17 + 10 + 7 + 4 = 67$

$\sum f_i x_i = 102 + 160 + 345 + 442 + 290 + 224 + 140 = 1703$

$$\bar{x} = \frac{1703}{67} = \boxed{25.42}$$

x_i: 5,10,15,20,25,30,35,40; f_i: 2,3,5,6,4,3,2,1 (N-W'25)

$N = 26$, $\sum fx = 10 + 30 + 75 + 120 + 100 + 90 + 70 + 40 = 535$,

$$\bar{x} = \frac{535}{26} = \boxed{20.58}$$

5.4 Mean Deviation (MD)

Average of absolute deviations from the mean

Mean Deviation — Formulas

Ungrouped Data

$$\text{MD} = \frac{\sum |x_i - \bar{x}|}{n}$$

Grouped / Frequency Data

$$\text{MD} = \frac{\sum f_i |x_i - \bar{x}|}{N} \quad \text{where } N = \sum f_i$$

Steps

1. Compute $\bar{x}$ first 2. Find $|x_i - \bar{x}|$ for each value 3. Multiply by f_i (grouped) 4. Divide by N

MD (Ungrouped) — GTU W'22

Data: 4,6,2,4,5,4,4,5,3,4 (3 marks)

$$n = 10, \;\; \bar{x} = \frac{41}{10} = 4.1$$

x_i	4	6	2	4	5	4	4	5	3	4
$\lvert x_i - 4.1 \rvert$	0.1	1.9	2.1	0.1	0.9	0.1	0.1	0.9	1.1	0.1

$$\sum \lvert x_i - \bar{x} \rvert = 7.4 \qquad \text{MD} = \frac{7.4}{10} = \boxed{0.74}$$

MD (Ungrouped) — GTU Su'25 & N-W'25

30,35,39,41,36,37,43,40,32 — Su'25 (3 marks)

$n = 9$, $\bar{x} = \frac{333}{9} = 37$

$|d_i|$: 7,2,2,4,1,0,6,3,5 $\Rightarrow \sum = 30$ $\quad$ $\text{MD} = \dfrac{30}{9} = \boxed{3.33}$

4,6,8,10,12,14 — N-W'25 (3 marks)

$n = 6$, $\bar{x} = \frac{54}{6} = 9$

$|d_i|$: 5,3,1,1,3,5 $\Rightarrow \sum = 18$ $\quad$ $\text{MD} = \dfrac{18}{6} = \boxed{3}$

MD (Grouped) — GTU Su'24

x_i: 3,4,5,6,7,8; f_i: 1,3,7,5,2,2 (3 marks)

$N = 20$, $\sum fx = 1\cdot3+3\cdot4+7\cdot5+5\cdot6+2\cdot7+2\cdot8 = 3+12+35+30+14+16 = 110$, $\bar{x} = 5.5$

x_i	f_i	$\lvert x_i - 5.5\rvert$	$f\lvert x_i - 5.5\rvert$
3	1	2.5	2.5
4	3	1.5	4.5
5	7	0.5	3.5
6	5	0.5	2.5
7	2	1.5	3.0
8	2	2.5	5.0
20			**21.0**

$$\text{MD} = \frac{21}{20} = \boxed{1.05}$$

MD (Grouped Classes) — GTU W'23 / Su'25

Class 30–100, *f*: 3,7,12,15,8,3,2 (3 marks)

Class	x	f	fx	$\lvert x-\bar{x}\rvert$	$f\lvert x-\bar{x}\rvert$
30–40	35	3	105	29	87
40–50	45	7	315	19	133
50–60	55	12	660	9	108
60–70	65	15	975	1	15
70–80	75	8	600	11	88
80–90	85	3	255	21	63
90–100	95	2	190	31	62
		50	**3100**		**556**

$\bar{x} = \frac{3100}{50} = 62 \qquad \text{MD} = \frac{556}{50} = \boxed{11.12}$

MD (Grouped) — GTU Su'25 & W'23

Class 0–50, f: 5,8,15,16,6 — Su'25 (3 marks)

x: 5,15,25,35,45; $N = 50$; $\sum fx = 25 + 120 + 375 + 560 + 270 = 1350$; $\bar{x} = 27$

$f|x - 27|$: $5 \cdot 22 + 8 \cdot 12 + 15 \cdot 2 + 16 \cdot 8 + 6 \cdot 18 = 110 + 96 + 30 + 128 + 108 = 472$

$$\text{MD} = \frac{472}{50} = \boxed{9.44}$$

x_i: 4,8,11,17,20,24,32; f_i: 3,5,9,5,4,3,1 — W'23/W'24 (3 marks)

$N = 30$; $\sum fx = 12 + 40 + 99 + 85 + 80 + 72 + 32 = 420$; $\bar{x} = 14$

$f|x-14|$: $3\cdot10+5\cdot6+9\cdot3+5\cdot3+4\cdot6+3\cdot10+1\cdot18 = 30+30+27+15+24+30+18 = 174$

$$\text{MD} = \frac{174}{30} = \boxed{5.8}$$

5.5 Standard Deviation (σ)

Root mean square deviation from the mean

Standard Deviation — Formulas

Ungrouped Data

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n}} \qquad \text{or} \qquad \sigma = \sqrt{\frac{\sum x_i^2}{n} - \bar{x}^2}$$

Grouped / Frequency Data

$$\sigma = \sqrt{\frac{\sum f_i(x_i - \bar{x})^2}{N}} \qquad \text{or} \qquad \sigma = \sqrt{\frac{\sum f_i x_i^2}{N} - \bar{x}^2}$$

Variance

$\text{Var} = \sigma^2$

SD (Ungrouped) — GTU W'23

Data: 6,7,10,12,13,4,8,12 (3 marks)

$$n = 8,\ \bar{x} = \frac{72}{8} = 9$$

$(x_i - 9)^2$: $9, 4, 1, 9, 16, 25, 1, 9$ $\quad \sum(x_i - \bar{x})^2 = 74$

$$\sigma = \sqrt{\frac{74}{8}} = \sqrt{9.25} = \boxed{3.04}$$

Data: 10,12,13,16,20,25 N-W'25 (3 marks)

$n = 6,\ \bar{x} = 16;\quad (d_i)^2$: $36, 16, 9, 0, 16, 81;\ \sum = 158$

$$\sigma = \sqrt{\frac{158}{6}} = \sqrt{26.33} = \boxed{5.13}$$

SD (Ungrouped) — GTU Su'24 / W'24

120,132,148,136,142,140,165,153 (4 marks)

$n = 8$, $\bar{x} = \dfrac{1136}{8} = 142$

x	$d = x - 142$	d^2
120	-22	484
132	-10	100
148	6	36
136	-6	36
142	0	0
140	-2	4
165	23	529
153	11	121
		1310

$$\sigma = \sqrt{\frac{1310}{8}} = \sqrt{163.75} = \boxed{12.80}$$

SD (Grouped Freq Table) — GTU W'22 / W'24

x_i: 4,8,11,17,20,24,32; f_i: 3,5,9,5,4,3,1 (3 marks)

$N = 30$; $\bar{x} = \frac{420}{30} = 14$

x_i	f_i	$d_i = x_i - 14$	d_i^2	$f_i d_i^2$
4	3	-10	100	300
8	5	-6	36	180
11	9	-3	9	81
17	5	3	9	45
20	4	6	36	144
24	3	10	100	300
32	1	18	324	324
30				**1374**

$$\sigma = \sqrt{\frac{1374}{30}} = \sqrt{45.8} = \boxed{6.77}$$

SD (Grouped Classes) — GTU W'24 (4 marks)

Class 0–100 (width 20), f: 12,38,42,23,5

Class	x	f	fx	$d = x - \bar{x}$	fd^2
0–20	10	12	120	-46	25392
20–40	30	38	1140	-26	25688
40–60	50	42	2100	-6	1512
60–80	70	23	1610	14	4508
80–100	90	5	450	34	5780
		120	**5420**		**62880**

$\bar{x} = \frac{5420}{120} = 45.17 \quad \sigma = \sqrt{\frac{62880}{120}} = \sqrt{524} = \boxed{22.89}$

SD (Grouped Freq Table) — GTU N-W'25

x_i: 2,4,6,8,10,12; f_i: 1,2,3,3,2,1 (3 marks)

$N = 12$; $\bar{x} = \frac{72}{12} = 7$

$(d_i)^2$: 25,9,1,1,9,25; fd^2: 25,18,3,3,18,25; $\sum fd^2 = 92$

$$\sigma = \sqrt{\frac{92}{12}} = \sqrt{7.67} = \boxed{2.77}$$

5.6 Coefficient of Variation (CV)

Relative measure of dispersion

Coefficient of Variation

Formula

$$CV = \frac{\sigma}{\bar{x}} \times 100\%$$

Interpretation

- Lower CV $\Rightarrow$ more consistent / less variable data
- Higher CV $\Rightarrow$ more spread out data
- Used to compare variability of two different datasets

Quick Example

If $\bar{x} = 50$ and $\sigma = 10$, then $CV = \dfrac{10}{50} \times 100 = 20\%$

5.7 GTU PYQ Practice

All exam-style questions from W'22 – N-W'25

PYQ — 3-Mark Questions (Mean & MD)

Try These

1. Find mean: x_i=92,93,97,98,102,104; f_i=3,2,3,2,6,4 *[W'22]*
2. Mean of 15,7,6,a,3 is 7; find a. *[W'22]*
3. MD about mean: 4,6,2,4,5,4,4,5,3,4 *[W'22]*
4. Find MD: x_i=3,4,5,6,7,8; f_i=1,3,7,5,2,2 *[Su'24]*
5. Grouped MD: class 30–100, f=3,7,12,15,8,3,2 *[W'23]*
6. Missing f: mean=19, x_i=6,10,14,18,24,28,30 *[Su'23]*

PYQ — 3-mark (SD) & 4-mark Questions

3-Mark (SD)

1. SD of 6,7,10,12,13,4,8,12 *[W'23]*
2. SD: x_i=4,8,11,17,20,24,32; f_i=3,5,9,5,4,3,1 *[W'22/W'24]*
3. SD of 10,12,13,16,20,25 *[N-W'25]*
4. SD: x_i=2,4,6,8,10,12; f_i=1,2,3,3,2,1 *[N-W'25]*

4-Mark

1. Mean (step deviation): class 21–50, f=8,10,24,30,12,16 *[W'22]*
2. SD of 10,15,7,19,9,21,23,25,26,30 *[Su'23]*
3. SD: 120,132,148,136,142,140,165,153 *[Su'24/W'24]*
4. SD (grouped 0–100): f=12,38,42,23,5 *[W'24]*

Unit 5 — Formula Summary

Central Tendency

Direct: $\bar{x} = \dfrac{\sum fx}{N}$

Assumed Mean: $\bar{x} = A + \dfrac{\sum fd}{N}$

Step Dev: $\bar{x} = A + h \cdot \dfrac{\sum fu}{N}$

Combined: $\bar{x}_{12} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$

Dispersion

Range $= x_{\max} - x_{\min}$

$\text{MD} = \dfrac{\sum f|x - \bar{x}|}{N}$

$\sigma = \sqrt{\dfrac{\sum f(x - \bar{x})^2}{N}}$

$\text{CV} = \dfrac{\sigma}{\bar{x}} \times 100\%$

Exam Weightage (Unit 5)

MCQ (1 mark) $\times$ 2–3 | Short (3 marks) $\times$ 2–3 | Long (4 marks) $\times$ 1

5.8 Correlation

Karl Pearson's Coefficient

Correlation — Definition and Formula

What is Correlation?

Measures the **strength and direction** of linear relationship between two variables x and y.

Karl Pearson's Coefficient r

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \cdot \sum(y-\bar{y})^2}}$$

Shortcut: $r = \dfrac{n\sum xy - \sum x \sum y}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$

Interpretation

$r = +1$: Perfect positive $\quad r = -1$: Perfect negative $\quad r = 0$: No correlation
$|r| > 0.7$: Strong $\quad 0.4 < |r| \leq 0.7$: Moderate $\quad |r| \leq 0.4$: Weak

Correlation Example 1 (GTU W'22, 4-mark)

Find r for: x: 1,2,3,4,5; y: 2,4,5,4,5

x	y	x^2	y^2	xy
1	2	1	4	2
2	4	4	16	8
3	5	9	25	15
4	4	16	16	16
5	5	25	25	25
15	20	55	86	66

Calculation

$$n = 5;\ r = \frac{5(66) - 15 \cdot 20}{\sqrt{[5(55) - 225][5(86) - 400]}} = \frac{330 - 300}{\sqrt{50 \cdot 30}} = \frac{30}{\sqrt{1500}} \approx \mathbf{0.775}$$

Correlation Example 2 (GTU Su'24, W'23)

Su'24: x: 6,2,10,4,8; y: 9,11,5,8,7

$n = 5$; $\sum x = 30$, $\sum y = 40$, $\sum x^2 = 220$, $\sum y^2 = 340$, $\sum xy = 225$

$$r = \frac{5(225) - 30 \cdot 40}{\sqrt{[5(220) - 900][5(340) - 1600]}} = \frac{-75}{\sqrt{200 \cdot 100}} = \frac{-75}{\sqrt{20000}} \approx \mathbf{-0.530}$$

Moderate negative correlation

W'23: x: 1,3,5,7,9; y: 2,6,8,6,10 — check

$n = 5$; $\sum x = 25$, $\sum y = 32$, $\sum x^2 = 165$, $\sum y^2 = 240$, $\sum xy = 190$

$$r = \frac{5(190) - 25 \cdot 32}{\sqrt{[5(165) - 625][5(240) - 1024]}} = \frac{150}{\sqrt{200 \cdot 176}} \approx \mathbf{0.796}$$

Strong positive correlation

5.9 Regression Lines

Lines of Best Fit

Regression — Lines of Best Fit

Two Regression Lines

y **on** x**:** $y - \bar{y} = b_{yx}(x - \bar{x})$ where $b_{yx} = r\dfrac{\sigma_y}{\sigma_x}$

x **on** y**:** $x - \bar{x} = b_{xy}(y - \bar{y})$ where $b_{xy} = r\dfrac{\sigma_x}{\sigma_y}$

Shortcut: $b_{yx} = \dfrac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2}$

Key Property

Both regression lines pass through $(\bar{x}, \bar{y})$

$r^2 = b_{yx} \cdot b_{xy} \Rightarrow r = \pm\sqrt{b_{yx} \cdot b_{xy}}$

Regression Example (GTU W'24, 4-mark)

x: 2,4,6,8,10; y: 5,7,9,8,11 — find both regression lines

$n = 5$; $\sum x = 30$, $\sum y = 40$, $\sum x^2 = 220$, $\sum y^2 = 340$, $\sum xy = 258$
$\bar{x} = 6$, $\bar{y} = 8$

y on x

$$b_{yx} = \frac{5(258) - 30 \cdot 40}{5(220) - 900} = \frac{90}{200} = 0.45$$

$$y - 8 = 0.45(x - 6) \Rightarrow y = 0.45x + 5.3$$

x on y

$$b_{xy} = \frac{5(258) - 30 \cdot 40}{5(340) - 1600} = \frac{90}{100} = 0.9$$

$$x - 6 = 0.9(y - 8) \Rightarrow x = 0.9y - 1.2$$

Regression: Finding r from Regression Lines (GTU W'23)

W'23: given $4x - 5y + 33 = 0$ and $20x - 9y - 107 = 0$

Identify which is y on x (steeper slope $= b_{yx}$):
From $4x - 5y + 33 = 0$: $y = \frac{4}{5}x + \frac{33}{5} \Rightarrow b_{yx} = 0.8$
From $20x - 9y - 107 = 0$: $x = \frac{9}{20}y + \frac{107}{20} \Rightarrow b_{xy} = \frac{9}{20} = 0.45$
$r = \sqrt{0.8 \times 0.45} = \sqrt{0.36} = \mathbf{0.6}$
Mean: solve simultaneously: $\bar{x} = 13$, $\bar{y} = 17$

PYQ — Correlation (4-mark practice set)

All PYQ Correlation Questions

1. x: 2,4,6,8,10; y: 5,7,9,8,11 — find r [Su'23]
2. x: 1,2,3,4,5,6; y: 6,4,3,5,4,2 — find r [W'22,W'24]
3. x: 9,8,7,6,5; y: 10,9,8,7,11 — r? [Su'25]
4. x: 12,10,8,6,4; y: 20,22,24,26,28 — r? [N-W'25]
5. Given $b_{yx} = 0.6$, $b_{xy} = 0.4$: find r [MCQ type]

PYQ Regression

1. Find both regression lines for x: 1,2,3,4,5; y: 3,4,4,5,4 [W'22]
2. Lines: $3x - 4y + 8 = 0$; $4x - 3y - 1 = 0$; find $\bar{x}, \bar{y}, r$ [Su'24]

SD — Shortcut Formula Examples

Shortcut SD Formula

$$\sigma = \sqrt{\frac{\sum fx^2}{N} - \left(\frac{\sum fx}{N}\right)^2} \quad \text{(avoid big decimals!)}$$

GTU Su'25: x: 3,8,13,18,23; f: 7,10,15,10,8 ($N = 50$)

$\sum fx = 3(7) + 8(10) + 13(15) + 18(10) + 23(8) = 21 + 80 + 195 + 180 + 184 = 660$
$\sum fx^2 = 63 + 640 + 2535 + 3240 + 4232 = 10710$
$\bar{x} = 660/50 = 13.2$; $\sigma = \sqrt{10710/50 - 13.2^2} = \sqrt{214.2 - 174.24} = \sqrt{39.96} \approx 6.32$

Combined SD & Two-Dataset Comparison (GTU N-W'25)

Which batch is more consistent?

Batch A: $\bar{x}_A = 40$, $\sigma_A = 6$ Batch B: $\bar{x}_B = 50$, $\sigma_B = 7$
$\text{CV}_A = \frac{6}{40} \times 100 = 15\%$ $\text{CV}_B = \frac{7}{50} \times 100 = 14\%$
$\text{CV}_B < \text{CV}_A \Rightarrow$ Batch B is **more consistent**

GTU W'24: Which dataset more variable?

Dataset 1: mean=25, $\sigma = 4$; $\text{CV}_1 = 16\%$
Dataset 2: mean=12, $\sigma = 3$; $\text{CV}_2 = 25\%$
Dataset 2 is **more variable**

Weighted Mean & Median (Quick Coverage)

Weighted Mean

$$\bar{x}_w = \frac{\sum w_i x_i}{\sum w_i}$$

Example: marks 70,80,60 with weights 3,4,2:

$$\bar{x}_w = \frac{3(70)+4(80)+2(60)}{9} = \frac{210+320+120}{9} = \frac{650}{9} \approx 72.2$$

Median (Grouped)

$$\text{Median} = L + \frac{\frac{N}{2} - F}{f} \times h$$

L=lower boundary of median class; F=cumulative freq before; f=freq of median class; h=class width

GTU MCQ Bank — Statistics (All Years)

1-Mark Questions

1. If CV of A$=12\%$, B$=18\%$, which is more consistent? **A**
2. SD of 5,5,5,5,5 is ______. **0**
3. If all values multiplied by k, new SD $=$ ______. $k\sigma$
4. $r=?$ if $b_{yx}=0.8$ and $b_{xy}=0.5$. $\sqrt{0.4}\approx 0.632$
5. Regression y on x: passes through $(\bar{x},\bar{y})$? **Yes**
6. Range of r: ______. $[-1,1]$
7. Mean deviation is least about ______. **median**
8. σ^2 is called ______. **variance**

PYQ 4-mark — Full Correlation & Regression Set

Su'24: Find regression line y on x

x: 5,10,15,20,25; y: 16,19,23,26,30; $n = 5$

$\sum x = 75$, $\sum y = 114$, $\sum x^2 = 1375$, $\sum xy = 1890$

$b_{yx} = \frac{5(1890) - 75 \cdot 114}{5(1375) - 5625} = \frac{810}{1250} = 0.648$

$\bar{x} = 15$, $\bar{y} = 22.8$; Line: $y - 22.8 = 0.648(x - 15) \Rightarrow y = 0.648x + 13.08$

N-W'25: Find r from: x: 10,20,30,40,50; y: 40,30,20,10,5

$\sum x = 150$, $\sum y = 105$, $\sum x^2 = 5500$, $\sum y^2 = 3125$, $\sum xy = 2750$

$r = \frac{5(2750) - 150 \cdot 105}{\sqrt{[5(5500) - 22500][5(3125) - 11025]}} = \frac{-2250}{\sqrt{5000 \cdot 4600}} \approx -0.936$ (strong negative)

Standard Deviation — Step Deviation Method

Formula for Grouped Data (Step Deviation)

$$\sigma = h\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2} \text{ where } u = \frac{x-A}{h}$$

GTU Su'25 (4-mark): Class 0–50, f: 5,8,15,12,10 ($N = 50$, $A = 25$, $h = 10$)

Class	x	u	fu	fu^2
0–10	5	-2	-10	20
10–20	15	-1	-8	8
20–30	25	0	0	0
30–40	35	1	12	12
40–50	45	2	20	40
			14	80

$\sigma = 10\sqrt{80/50 - (14/50)^2} = 10\sqrt{1.6 - 0.0784} = 10\sqrt{1.5216} \approx \mathbf{12.34}$

Frequency Distribution — Missing Frequency Problems

GTU W'24: Mean = 50, find missing f_3

x:	10	30	50	70	90
f:	17	f_2	32	f_4	19

Given: $N = 120$, mean $= 50$; $f_2 + f_4 = 52$; $\sum fx =$ mean $\times N = 6000$

$170 + 30f_2 + 1600 + 70f_4 + 1710 = 6000 \Rightarrow 30f_2 + 70f_4 = 2520$

Solve with $f_2 + f_4 = 52$: $f_2 = 28$, $f_4 = 24$

All Key Formulas at a Glance

Mean (direct)	$\bar{x} = \frac{\sum fx}{N}$
Mean (assumed)	$\bar{x} = A + \frac{\sum fd}{N}$
Mean (step dev)	$\bar{x} = A + h \cdot \frac{\sum fu}{N}$
MD	$\frac{\sum f\lvert x-\bar{x}\rvert}{N}$
SD (direct)	$\sqrt{\frac{\sum f(x-\bar{x})^2}{N}}$
SD (shortcut)	$\sqrt{\frac{\sum fx^2}{N} - \bar{x}^2}$
SD (step dev)	$h\sqrt{\frac{\sum fu^2}{N} - \left(\frac{\sum fu}{N}\right)^2}$
CV	$\frac{\sigma}{\bar{x}} \times 100\%$
Pearson r	$\frac{n\sum xy-\sum x\sum y}{\sqrt{[\ldots][\ldots]}}$
b_{yx}	$\frac{n\sum xy-\sum x\sum y}{n\sum x^2-(\sum x)^2}$
r from b's	$r = \pm\sqrt{b_{yx} \cdot b_{xy}}$

GTU Frequency Guide — Unit 5 All Topics

Marks Distribution Across Papers

Topic	Marks	Frequency
Mean (direct/step dev)	3–4	Every paper
Mean Deviation	3	1–2 per paper
Standard Deviation	3–4	Every paper
CV comparison	3	W'24, Su'25, N-W'25
Missing frequency	3–4	W'22, Su'23, W'24
Correlation (Pearson r)	4	W'22, Su'24, W'23, W'24
Regression lines	4	Su'24, W'23, N-W'25
MCQs (all formulae)	1 each	2–3 per paper

Unit 5 & Course Complete!

Applied Mathematics — DI02000011

All 5 units covered. Best of luck!

SD Properties — Important Results

Effect of Change on SD

Add constant k	SD unchanged (σ stays same)
Multiply by k	New SD $= k\sigma$
All values equal	SD $= 0$
First n **naturals**	$\sigma = \sqrt{\frac{n^2-1}{12}}$

Variance vs SD

Variance $= \sigma^2$; SD $= \sigma = \sqrt{\text{Variance}}$

Variance of first 6 naturals $= \frac{6^2-1}{12} = \frac{35}{12} \approx 2.917$; SD$\approx 1.708$

Mixed PYQ Set — 1-mark & 3-mark (All Years)

3-Mark: Quick Calculations

1. Find mean: 15,20,25,30,35 — $\bar{x} = 25$
2. SD of 2,4,4,4,5,5,7,9 — $\sigma = 2$ [$W'22$]
3. If $\bar{x} = 30$, $\sigma = 6$: find CV — CV $= 20\%$
4. $r = 0.8$, $\sigma_x = 3$, $\sigma_y = 4$: find b_{yx} — $b_{yx} = \frac{0.8\times 4}{3} \approx 1.067$
5. MD about mean: 4,7,2,9,3 — $\bar{x} = 5$; MD$= 2.4$

Rank Correlation (Spearman) — GTU N-W'25

Spearman's Rank Correlation

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \text{ where } d_i = R(x_i) - R(y_i)$$

N-W'25: x: 56,75,45,71,62,64,58,80,76; y: 66,70,40,60,65,56,59,77,67

Rank x: 2,7,1,6,4,5,3,9,8 Rank y: 5,8,1,3,4,2,7,9,6
d: $-3, -1, 0, 3, 0, 3, -4, 0, 2$; $\sum d^2 = 9 + 1 + 0 + 9 + 0 + 9 + 16 + 0 + 4 = 48$
$r_s = 1 - \frac{6(48)}{9(80)} = 1 - \frac{288}{720} = 1 - 0.4 = \mathbf{0.6}$

Exam Strategy — Unit 5

How to Score Full Marks in Statistics

- **MCQ:** Know $r \in [-1, 1]$; lower CV = more consistent; SD of constant data = 0
- **3-mark:** Show table (x, f, fx, etc.) — don't skip steps
- **4-mark:** For correlation, show all five columns (x, y, x^2, y^2, xy) + totals
- **Regression:** compute both b_{yx} & b_{xy} even if only one line asked
- **SD:** Use shortcut formula for grouped data to save time

Time Allocation (Unit 5 in 80-mark paper)

MCQ: 5 min | Short: 10 min each | Long: 15 min