FREE EDITION · NOTES + 3 SAMPLE MCQS

CUET · MATHEMATICS · CLASS XI · CODE 319

Statistics

CUET unit: Statistics

By **UniDrill** · NCERT-grounded study material

WWW.UNIDRILL.IN

📌 Snapshot

- Measures of central tendency (mean, median, mode) alone cannot describe a data set — two series with the same mean can have very different scatter (e.g. batsmen A and B, both with mean and median 53 but vastly different consistency).
- The four measures of dispersion are range, quartile deviation, mean deviation, and standard deviation (quartile deviation is named but not studied here).
- Builds mean deviation about mean and median for ungrouped, discrete-frequency, and continuous-frequency data, using absolute deviations.
- Develops variance σ^2 and standard deviation σ as the mathematically superior dispersion measures (using squared deviations) and derives an alternative computational formula and a step-deviation short-cut.
- CUET regularly tests numerical computation of MD, variance and SD plus conceptual results from the Miscellaneous Examples (effect of multiplying/ adding a constant, correcting wrong observations).

📖 Detailed Notes

2.1 Core concepts

- Measures of central tendency (mean, median, mode) only locate the centre of data; they say nothing about how scattered the data are around that centre — hence the need for a single number called a measure of dispersion (NCERT §13.1, p. 258). Two distributions can share mean and median yet have very different spreads.
- The four measures of dispersion are Range, Quartile deviation, Mean deviation, and Standard deviation; Quartile deviation is excluded from study (NCERT §13.2, p. 259). Range is the simplest; SD is the most informative.
- **Range** of a series = Maximum value – Minimum value; it gives a rough idea of scatter but does not measure dispersion about a central tendency (NCERT §13.3, p. 259). Range is unstable: a single outlier can inflate it dramatically.
- The deviation of an observation x from a fixed value a is x – a; the sum of deviations from the mean is always zero, so the mean of signed deviations is useless as a dispersion measure (NCERT §13.4, p. 259). This is a key consequence of the definition $\bar{x} = (\Sigma x_i)/n$.

- To overcome cancellation of signs, we take **absolute** deviations; the mean of absolute deviations from a central value a is the mean deviation M.D.(a) = $\Sigma|x_i - a| / n$ (NCERT §13.4, p. 260). Mean deviation is intuitive but mathematically awkward (the absolute-value function is not smooth).
- For ungrouped data: M.D.($\bar{x}$) = $(1/n)\ \Sigma|x_i - \bar{x}|$; M.D.(M) = $(1/n)\ \Sigma|x_i - M|$, where M is the median (NCERT §13.4.1, p. 260). M.D. about median is minimum: among all choices of central value a, M.D.(a) is least when a is the median.
- For a discrete frequency distribution with $N = \Sigma f_i$: M.D.($\bar{x}$) = $(1/N)\ \Sigma f_i|x_i - \bar{x}|$, and median is found by locating the observation whose cumulative frequency is equal to or just greater than N/2 (NCERT §13.4.2, pp. 262–264).
- For a continuous frequency distribution, mid-points x_i of each class are used; the median class is the class whose c.f. is just greater than or equal to N/2, and Median $= l + ((N/2 - C)/f) \times h$, where l, f, h are the lower limit, frequency and width of the median class and C is the c.f. of the preceding class (NCERT §13.4.2, p. 268).
- A step-deviation short-cut for mean is $d_i = (x_i - a)/h$, giving $\bar{x} = a + h\cdot(\Sigma f_i d_i)/N$ (NCERT §13.4.2, p. 267). Choosing a near the centre of the data minimises arithmetic.
- Limitations of mean deviation — it uses absolute values so cannot be subjected to further algebraic treatment, and M.D. about median is unreliable when variability is high (NCERT §13.4.3, p. 271). This motivates moving to squared deviations.
- To avoid the absolute-value problem, square the deviations instead. The simple sum $\Sigma(x_i - \bar{x})^2$ is not a good measure (it grows with n), so we take its mean: **variance $\sigma^2 = (1/n)\ \Sigma(x_i - \bar{x})^2$** for ungrouped data (NCERT §13.5, pp. 271–274).
- Because variance has the square of the observations, the proper dispersion measure is the positive square root, **standard deviation $\sigma = \sqrt{[(1/n)\ \Sigma(x_i - \bar{x})^2]}$** (NCERT §13.5.1, p. 274). SD is in the same units as the data, making it directly interpretable.
- For a discrete frequency distribution: $\sigma = \sqrt{[(1/N)\ \Sigma f_i(x_i - \bar{x})^2]}$ (NCERT §13.5.2, p. 275).
- For a continuous frequency distribution (replacing each class by its mid-point): $\sigma = \sqrt{[(1/N)\ \Sigma f_i(x_i - \bar{x})^2]}$, and an algebraically equivalent computational form is $\sigma = (1/N)\sqrt{[N\ \Sigma f_i x_i^2 - (\Sigma f_i x_i)^2]}$ (NCERT §13.5.3, pp. 276–277). The alternative form avoids computing $\bar{x}$ first.
- **Short-cut (step-deviation) method:** with $y_i = (x_i - A)/h$, $\bar{x} = A + h\cdot\bar{y}$ and $\sigma_x = h\cdot\sigma_y$, giving $\sigma^2 = (h^2/N^2)[N\ \Sigma f_i y_i^2 - (\Sigma f_i y_i)^2]$ (NCERT §13.5.4, pp. 279–281). The factor h^2 appears because variance is squared in unit; SD scales by |h|.
- Key results from Miscellaneous Examples — (a) multiplying each observation by k multiplies the variance by k^2 (so SD by |k|) (Example 13, p. 282); (b) adding a constant a to every observation leaves the variance unchanged (Example 15, p. 284); (c) given mean and SD with a wrong observation, the correct mean and SD are found by adjusting Σx_i and Σx_i^2 (Example 16, pp. 284–285). All three results recur on CUET papers.

- The coefficient of variation CV = $(\sigma/\bar{x})\cdot 100$ compares dispersions of data sets with different units or magnitudes; lower CV means greater consistency. CV is dimensionless, allowing comparison across units (e.g., height in cm vs weight in kg).
- Variance divides by n (not n – 1) — this is the "population" variance, appropriate when the data set is treated as exhaustive. The unbiased estimator using n – 1 is part of higher statistics, not Class XI.
- Worked tip for the alternative SD formula: build a small table of x_i^2, multiply by f_i, sum column-wise; one pass yields both $\Sigma f_i x_i$ and $\Sigma f_i x_i^2$. This is more efficient than computing each $(x_i - \bar{x})^2$ individually.
- Skewness comparison via Misc. Example 14 (p. 283): when one data set has a larger SD than another, the first is more dispersed; this is the basis of "consistency comparison" in cricket-batsman / quality-control problems on CUET.
- For comparing two series where the means are equal, the series with smaller SD is more consistent; where the means differ, the coefficient of variation provides a meaningful comparison.
- Practical computational warning: in step-deviation, choose A as one of the x_i (often the middle observation) and h as the common difference of class widths so that the y_i become small integers like –3, –2, ⋯, 3. This dramatically simplifies arithmetic and reduces transcription errors.
- Moving from range → MD → SD trades more computation for better statistical behaviour (algebraic tractability, sensitivity to all observations, units consistency, invariance under translation). This is why higher statistics — regression, hypothesis testing, ANOVA — is built on variance and SD, not on range or mean deviation.

2.2 Definitions to memorise

Term	Definition	Page
Measure of dispersion	Single number describing scatter	258
Range	Max – Min	259
Deviation of x from a	x – a	259
Sum of deviations from mean	Always 0	259
Mean deviation M.D.(a)	(1/n) Σ	$x_i - a$
M.D. about mean (ungrouped)	(1/n) Σ	$x_i - \bar{x}$
M.D. about median (ungrouped)	(1/n) Σ	$x_i - M$
M.D. (discrete frequency)	(1/N) Σf_i	$x_i - \bar{x}$
Median class	C.f. just ⩾ N/2	268
Median (continuous)	l + ((N/2 – C)/f)·h	268
Variance σ^2 (ungrouped)	$(1/n)\ \Sigma(x_i - \bar{x})^2$	274

Term	Definition	Page
Standard deviation σ	$\sqrt{\text{variance}}$	274
SD (discrete frequency)	$\sqrt{[(1/N)\ \Sigma f_i(x_i - \bar{x})^2]}$	275
Alternative SD	$(1/N)\ \sqrt{[N\ \Sigma f_i x_i^2 - (\Sigma f_i x_i)^2]}$	277
Step-deviation y_i	$(x_i - A)/h$	280
Step-deviation mean	$\bar{x} = A + h \cdot \bar{y}$	280
Step-deviation SD	$\sigma_x = h \cdot \sigma_y$	280
Effect of adding constant	Variance unchanged	284
Effect of multiplying by k	Variance × k^2	282
Correction for wrong observation	Adjust Σx_i and Σx_i^2	285
Coefficient of variation	$(\sigma/\bar{x}) \cdot 100\ \%$	285
Cumulative frequency C	Sum up to a class	268
$N = \Sigma f_i$	Total frequency	263
Class width h	Upper − lower limit	267
Mid-point of class	(lower + upper)/2	267
Assumed mean A	Reference value in step-deviation	280

2.3 Diagrams / processes to remember

- **Fig 13.1 & 13.2 (p. 258):** dot plots of batsman A's and batsman B's scores on a number line — visually shows that despite identical mean and median, A's data are scattered while B's cluster near the centre. Underscores why dispersion is needed.
- **Fig 13.3 (p. 267):** number line showing shifting of origin to assumed mean A in the step-deviation method.
- **Fig 13.4 (p. 267):** number line showing change of scale (division by common factor h) — together explain why $d_i = (x_i - A)/h$ works.
- **Fig 13.5 & 13.6 (p. 273):** dot plots for set A (6 observations: 5, 15, ⋯, 55) and set B (31 observations from 15 to 45) — geometrical proof that $\Sigma(x_i - \bar{x})^2$ alone is misleading and one must take its mean.
- **Tables 13.1 – 13.11 (pp. 263–280):** standardised tabular layouts for computing MD and variance — students should recall the column structure (x_i, f_i, $f_i x_i$, $|x_i - \bar{x}|$, $f_i|x_i - \bar{x}|$, $(x_i - \bar{x})^2$, $f_i(x_i - \bar{x})^2$, and for short-cut: y_i, y_i^2, $f_i y_i$, $f_i y_i^2$).
- **Process — mean deviation about mean (ungrouped):** (i) compute $\bar{x} = (\Sigma x_i)/n$; (ii) compute each $|x_i - \bar{x}|$; (iii) sum and divide by n.
- **Process — mean deviation about median (ungrouped):** (i) arrange observations in ascending order; (ii) identify median (middle term for odd n; mean of two middle terms for even n); (iii) compute $|x_i - M|$ for each; (iv) sum and divide by n.

- **Process — variance and SD (ungrouped):** (i) compute $\bar{x}$; (ii) compute each $(x_i - \bar{x})^2$; (iii) sum and divide by n for σ^2; (iv) take positive square root for σ.
- **Process — alternative computational SD:** (i) compute $\Sigma f_i x_i$ and $\Sigma f_i x_i^2$ directly; (ii) apply $\sigma = (1/N) \sqrt{[N \Sigma f_i x_i^2 - (\Sigma f_i x_i)^2]}$.
- **Process — step-deviation:** (i) choose convenient A and h; (ii) compute $y_i = (x_i - A)/h$; (iii) compute $\Sigma f_i y_i$ and $\Sigma f_i y_i^2$; (iv) $\bar{y} = \Sigma f_i y_i/N$; (v) $\bar{x} = A + h \cdot \bar{y}$; (vi) $\sigma = (h/N) \sqrt{[N \Sigma f_i y_i^2 - (\Sigma f_i y_i)^2]}$.

2.4 Common confusions / NTA trap points

- Forgetting the absolute value in mean deviation — $\Sigma(x_i - \bar{x})$ is always zero, so a student who skips the modulus will write the answer as 0 (NCERT §13.4, p. 259). NTA frequently sets a distractor of 0 for this reason.
- Confusing variance and standard deviation — variance is σ^2 (units squared); SD is the positive square root σ. Many students stop at variance and report it as SD.
- Mixing the divisor in grouped data — for mean deviation/variance of a frequency distribution the divisor is $N = \Sigma f_i$, not n (the number of distinct values).
- Applying the wrong formula for median class — the c.f. just greater than or equal to N/2 (not the c.f. just less than) identifies the median class, then Median = $l + ((N/2 - C)/f) \times h$.
- Effect of transformations — adding a constant does not change variance/SD, but multiplying every observation by k multiplies variance by k^2 (and SD by $|k|$). NTA mixes "added" and "multiplied" deliberately to trip students (Examples 13 & 15, pp. 282, 284).
- Step-deviation method is only a computational short-cut for mean and SD; the answer must come out the same as by the direct formula (NCERT §13.5.4, pp. 279–281).
- Treating range as a dispersion measure about the mean — range ignores the mean entirely.
- Reporting SD with a negative sign — SD is non-negative by definition.
- Forgetting that the mean shifts by the same constant when observations shift; failing to update $\bar{x}$ leads to wrong $(x_i - \bar{x})$.
- Mis-identifying the median class as "the one containing N/2" rather than "the one whose c.f. first reaches or exceeds N/2".
- Forgetting to multiply the deviation factor h when converting step-deviation SD back to original units ($\sigma_x = h \cdot \sigma_y$, not σ_y).
- Treating mean deviation as exchangeable with SD in numerical problems; they differ unless data are symmetric.

2.5 Key formulas & theorems

Formula	Statement	NCERT page
Range	Max – Min	259
Mean	x̄ = (Σxi)/n	260
Mean (frequency)	x̄ = (Σfixi)/N	263
Sum of deviations	Σ(xi – x̄) = 0	259
M.D. about mean	(1/n) Σ	xi – x̄
M.D. about median	(1/n) Σ	xi – M
M.D. (discrete freq.)	(1/N) Σfi	xi – x̄
Median (continuous)	l + ((N/2 – C)/f) h	268
Variance (ungrouped)	(1/n) Σ(xi – x̄)²	274
Standard deviation	√[(1/n) Σ(xi – x̄)²]	274
SD (discrete)	√[(1/N) Σfi(xi – x̄)²]	275
SD (continuous)	√[(1/N) Σfi(xi – x̄)²] with mid-points	277
Alternative SD	(1/N) √[N Σfixi² – (Σfixi)²]	277
Step-deviation y	yi = (xi – A)/h	280
Step-deviation x̄	A + h·ȳ	280
Step-deviation σ	h·σy	280
Step-deviation σ²	(h²/N²)[N Σfiyi² – (Σfiyi)²]	281
Constant shift effect	var(x + a) = var(x)	284
Scaling effect	var(kx) = k² var(x)	282
SD scaling	SD(kx) =	k
Coefficient of variation	(σ/x̄) · 100	285
Correction of mean	Σxi adjusted by Δ	285
Correction of variance	Σxi² adjusted by Δ²	285
Variance ⩾ 0	Always non-negative	274
Variance = 0 iff all equal	Constant data	274
Mean of deviations about a	(Σxi – na)/n	259

2.6 Solved examples (NCERT-grounded)

Example A (NCERT Example 1, p. 261). M.D. about mean of 6, 7, 10, 12, 13, 4, 8, 12.

Step 1 — mean: Σ = 72; n = 8; x̄ = 9. Step 2 — |deviations|: 3, 2, 1, 3, 4, 5, 1, 3; sum = 22. Step 3 — divide: M.D. = 22/8 = **2.75**.

Example B (NCERT Example 3, p. 262). M.D. about median of 3, 9, 5, 3, 12, 10, 18, 4, 7, 19, 21.

Step 1 — sort: 3, 3, 4, 5, 7, 9, 10, 12, 18, 19, 21; median = 6th term = 9. Step 2 — $|x_i - 9|$: 6, 6, 5, 4, 2, 0, 1, 3, 9, 10, 12; sum = 58. Step 3 — divide: M.D. = 58/11 ≈ **5.27**.

Example C (NCERT Example 8, p. 275). Variance of 6, 8, 10, 12, 14, 16, 18, 20, 22, 24.

Step 1 — mean: $\Sigma = 150$; $n = 10$; $\bar{x} = 15$. Step 2 — $\Sigma(x_i - \bar{x})^2$: 81 + 49 + 25 + 9 + 1 + 1 + 9 + 25 + 49 + 81 = 330. Step 3 — variance: 330/10 = **33** (SD = $\sqrt{33} \approx 5.74$).

Example D (NCERT Misc. Example 16, p. 285). Mean 40, SD 5.1 for n = 100. One value 50 was wrongly used instead of 40. Find corrected mean and SD.

Step 1 — correct Σx_i: $100 \cdot 40 = 4000$; corrected = 4000 − 50 + 40 = 3990; mean = 39.9. Step 2 — original Σx_i^2: var = $5.1^2 = 26.01 \Rightarrow \Sigma x_i^2/100 - 40^2 = 26.01 \Rightarrow \Sigma x_i^2 = 162601$; corrected = $162601 - 50^2 + 40^2 = 161701$. Step 3 — new variance and SD: var = $161701/100 - 39.9^2 = 1617.01 - 1592.01 = 25 \Rightarrow$ SD = **5**. Corrected: mean 39.9, SD 5.

Example E (NCERT Misc. Example 15, p. 284). Show that adding constant a to every observation does not change variance.

Step 1 — set $y_i = x_i + a$: $\bar{y} = \bar{x} + a$. Step 2 — compute deviations: $y_i - \bar{y} = (x_i + a) - (\bar{x} + a) = x_i - \bar{x}$. Step 3 — variance unchanged: $(1/n)\,\Sigma(y_i - \bar{y})^2 = (1/n)\,\Sigma(x_i - \bar{x})^2 = \mathrm{var}(x)$. **QED.**

🎯 Practice MCQs

📊 PYQ Alignment

CUET (UG) Mathematics typically draws 1–2 direct MCQs from this chapter every year, almost always numerical — computing mean deviation about mean/median for a small ungrouped set, computing variance/SD using either the direct formula or the alternative computational form $\sigma = (1/N)\sqrt{[N\,\Sigma f_i x_i^2 - (\Sigma f_i x_i)^2]}$, or applying the "wrong observation corrected" and "multiply/add a constant" results from the Miscellaneous Examples. Statement-based and assertion-reason items on the effect of linear transformations (add/multiply a constant) on variance are also common.