Unit 15 · Measure of Dispersion

STUDY NOTES

WHAT NEB'S GRADE 11 EXAM ASKS

- **CDC's grid (2078):** Statistics and Probability (Units 15 and 16 together): 6 of 75 marks, one 1-mark choice and one 5-mark short question; no 8-mark question from this area
- a dispersion question can ask for any of the mean, σ, the C.V. and Karl Pearson's coefficient of skewness, and what they say
- **MCQ (1 mark)**
  - the definition of variance or of S_k
  - a constant added; every value multiplied by 3
  - order of mean, median, mode in a skewed set; which set is more consistent
  - σ from Σx, Σx^2; S_k from mean, median, σ
- **Short (5 marks)**
  - mean, σ, C.V. of a table
  - whose sales are more consistent
  - correct mean and σ after a copying slip
  - $\Sigma(x-a)^2$ is least at $a=\overline{X}$
- **Long (school tests)**: mean, σ, C.V., median, mode, S_k both ways, and what they show

Key terms

Standard deviation σ: the square root of the mean of the squared deviations from the mean; in the data's unit.

Variance: σ^2; in the unit squared.

Coefficient of variation: σ as a percentage of the mean; no unit.

Coefficient of standard deviation: $\sigma/\overline{X}$.

Skewness: lack of symmetry of a distribution.

Positively skewed (to the right): a long tail of a few large values.

Negatively skewed (to the left): a long tail of a few small values.

Formulas

$\sigma=\sqrt{\Sigma(x-\overline{X})^2/n}$	individual values **Use:** definition
$\sigma^2=\Sigma x^2/n-(\Sigma x/n)^2$	grouped: $\Sigma fm^2/N-(\Sigma fm/N)^2$ **Use:** short cut
$\sigma=\sqrt{\Sigma fd'^2/N-(\Sigma fd'/N)^2}\times h$	$d'=(m-A)/h$, m mid-value, h width **Use:** step deviation
$\overline{X}=A+(\Sigma fd'/N)\times h$	same d' **Use:** mean
C.V. $=(\sigma/\overline{X})\times 100\%$	no unit **Use:** compare spreads
$S_k=(\overline{X}-M_o)/\sigma$	mode clear **Use:** Karl Pearson
$S_k=3(\overline{X}-M_d)/\sigma$	mode ill-defined; between -3 and 3 **Use:** Karl Pearson, median form

- from Grade 10: $M_d=L+(N/2-cf)/f\times h$; $M_o=L+(f_1-f_0)/(2f_1-f_0-f_2)\times h$

Standard deviation: why the short cuts work

Mean of the squares minus the square of the mean

1. $\sigma^2 = (1/n)\Sigma(x - \overline{X})^2$
2. open the square: $(1/n)\Sigma(x^2 - 2\overline{X}x + \overline{X}^2)$
3. term by term: $\Sigma x^2/n - 2\overline{X} \cdot \Sigma x/n + \overline{X}^2$
4. $\Sigma x/n = \overline{X} \rightarrow \Sigma x^2/n - \overline{X}^2$

Change of origin and scale: $u = (x - A)/h$, $x = A + hu$

1. $\overline{X} = A + h\overline{u}$
2. $x - \overline{X} = h(u - \overline{u})$
3. $\sigma_x{}^2 = h^2\sigma_u{}^2 \rightarrow \sigma_x = |h|\sigma_u$

- **add a constant**: mean moves, σ unchanged
- **multiply by** h: $\sigma \times |h|$, variance $\times h^2$
- calculator: σ_x ($\div n$) is ours; s_x ($\div n - 1$) is for a sample estimating a population
- report σ in the data's unit
- **least squares**: $\Sigma(x - a)^2$ is least at $a = \overline{X}$
- **a copying slip**: correct Σx and Σx^2 ($\Sigma x^2 = n(\sigma^2 + \overline{X}^2)$), then recompute

Coefficient of variation

- σ alone can't compare different sizes or different units $\rightarrow$ divide by the mean
- **smaller C.V.** $\rightarrow$ **more consistent** (uniform, stable); larger $\rightarrow$ more variable
- only for data from a true zero with a positive mean (rainfall, weight, money); not °C

Skewness

- **symmetric**: mean = median = mode
- **positive**: usually $M_o < M_d < \overline{X}$ (rainfall, incomes, waiting times)
- **negative**: usually $\overline{X} < M_d < M_o$ (marks in an easy test)
- the mean feels every value's size $\rightarrow$ dragged by the tail; the median counts only $\rightarrow$ moves less; the mode stays at the peak

Karl Pearson

- divide the gap by σ, so no unit
- $S_k > 0$ positive; < 0 negative; $= 0$ symmetric (as far as this measure sees)
- **mode ill-defined** (two peaks, no repeats, modal class at an end): use the median
- why: $M_o \approx 3M_d - 2\overline{X}$ (Pearson's empirical relation) $\rightarrow \overline{X} - M_o \approx 3(\overline{X} - M_d)$
- the two forms differ in value, usually agree in sign; use the one asked
- median form in $[-3, 3]$: mean and median never more than one σ apart

Kathmandu and Pokhara: monthly rain, WMO normals 1991–2020

	$\overline{X}$	σ	C.V.	M_d	S_k (median)
Kathmandu	125.25	130.22	104.0%	53	1.66
Pokhara	322.25	334.50	103.8%	150.5	1.54

- Pokhara's σ is 2.6 times, but so is its mean: C.V.s alike $\rightarrow$ **no more uneven**
- strongly positive skew: a typical month is the **median**, not the mean

For the engineering entrance (IOE)

- **quartiles**: like the median with $N/4$, $3N/4$: $Q_1 = L + (N/4 - cf)/f \times h$
- range = largest − smallest; quartile deviation $(Q_3 - Q_1)/2$
- **two groups**: $\overline{X} = (n_1\overline{X}_1 + n_2\overline{X}_2)/(n_1 + n_2)$
- $\sigma^2 = (n_1(\sigma_1{}^2 + d_1{}^2) + n_2(\sigma_2{}^2 + d_2{}^2))/(n_1 + n_2)$, $d_i = \overline{X}_i - \overline{X}$
  - larger than either group's when the means differ
- **1, 2, …, n**: mean $(n + 1)/2$; $\sigma = \sqrt{(n^2 - 1)/12}$
- **two numbers** a, b: $\sigma = |a - b|/2$

COMMON MISTAKES

✗ adding 5 to every value changes σ
✓ only the mean moves.

✗ multiplying by 3 triples the variance
✓ variance $\times 9$; $\sigma \times 3$.

✗ the larger σ, the less consistent
✓ compare C.V.s when the means differ.

✗ $\sigma^2 = \Sigma x^2/n - \overline{X}$
✓ minus $\overline{X}^2$.

✗ step deviation without $\times h$ at the end
✓ multiply σ by h.

✗ dividing by $n - 1$
✓ NEB's σ divides by n.

✗ positive skew: mean below the mode
✓ above it: the tail drags the mean.