Unit 15 · Measure of Dispersion

HARDER PRACTICE QUESTIONS · ANSWER KEY

Full marks 28

Give the marks for the points in the marking scheme; an answer in other words that says the same is right.

1 One more value

ANSWER HA · 6

Tests: σ^2 is a total of squared distances shared over n, so a new value changes the total, the count and the mean; rebuilding Σx, Σx^2 from summary figures. The tempting wrong answers: "unchanged" in a, and 50 ± 4 (one σ from the mean) in c.

$\Sigma x = 20 \times 50 = 1000$; $\Sigma x^2 = 20(\sigma^2 + \bar{X}^2) = 20(16 + 2500) = 50320$.

a. New: $\Sigma x = 1050$, $\Sigma x^2 = 52820$, $n = 21$, mean 50. $\sigma = \sqrt{\dfrac{52820}{21} - 2500} = \sqrt{\dfrac{320}{21}} \approx 3.90$.

b. The squared distances still total 320 (the new value adds 0), but now they are shared among 21 values instead of 20, so their mean, σ^2, is smaller.

c. Add x: $\dfrac{50320 + x^2}{21} - \left(\dfrac{1000 + x}{21}\right)^2 = 16$. Multiply by $21^2 = 441$:

$21(50320 + x^2) - (1000 + x)^2 = 7056$, that is $20x^2 - 2000x + 49664 = 0$, or $(x - 50)^2 = 16.8$.

$x = 50 \pm \sqrt{16.8} \approx 54.10$ or **45.90**. (Not 50 ± 4: the new value also pulls the mean towards itself, so it must lie a little further out than one σ.)

MARKING SCHEME

- ✓ $\Sigma x = 1000$ and $\Sigma x^2 = 50320$.
- ✓ a. $\sigma = \sqrt{320/21} \approx 3.90$.
- ✓ b. The same total of squared distances shared among more values.
- ✓ c. The equation for $\sigma^2 = 16$ with the new mean $\dfrac{1000 + x}{21}$.
- ✓ Reduced to $(x - 50)^2 = 16.8$.
- ✓ $x \approx 54.10$ or 45.90.

2 A fixed charge and a rate

ANSWER A · 6

Tests: change of origin and scale inside a real formula; which measures follow the shift and which don't. The tempting wrong answers: $\sigma = 150 + 10 \times 30 = 450$, and "the C.V. is the same".

a. Bill $= 150 + 10u$. Every bill is found this way, so the mean and the median follow: mean $= 150 + 10 \times 80 =$ **Rs 950**, median $= 150 + 10 \times 70 =$ **Rs 850**. The fixed charge shifts every bill alike and leaves the spread; the rate multiplies it: $\sigma = 10 \times 30 =$ **Rs 300**.

b. Units: $\dfrac{30}{80} \times 100\% = 37.5\%$. Bills: $\dfrac{300}{950} \times 100\% \approx 31.6\%$. The rate scales σ and the mean alike, but the fixed Rs 150 raises the mean without adding any spread, so the bills vary less compared with their size.

c. Units: $\dfrac{3(80 - 70)}{30} = 1$. Bills: $\dfrac{3(950 - 850)}{300} = 1$. **The same**: the shift cancels in mean minus median, and the scale cancels in the division by σ.

MARKING SCHEME

- ✓ Mean Rs 950 and median Rs 850.
- ✓ $\sigma =$ Rs 300 (the fixed charge doesn't add to it).
- ✓ The reason: a constant shifts all bills alike; the rate multiplies the spread.
- ✓ C.V. of the units 37.5% and of the bills about 31.6%.
- ✓ The reason: the fixed charge raises the mean but not σ.
- ✓ c. Both 1, unchanged.

Science Playground, Katunje, Bhaktapur
Book a class visit: https://science-playground.com/booking

3 A table that counts up

ANSWER A · 6

Tests: reading a cumulative ("less than") table, the shape NEB's model paper uses, before any formula. The tempting wrong answer: using 4, 12, 27, … as the frequencies (it gives $N = 179$ for 50 porters).

a. Frequencies are the differences: 4, 8, 15, 13, 6, 4 (total 50).

b. Mid-values 5, 15, …, 55; with $A = 35$, $h = 10$: $d' = -3, -2, -1, 0, 1, 2$;
$\Sigma fd' = -12 - 16 - 15 + 0 + 6 + 8 = -29$; $\Sigma fd'^2 = 36 + 32 + 15 + 0 + 6 + 16 = 105$.

$\overline{X} = 35 + \frac{-29}{50} \times 10 = 29.2$ (Rs 2,920). $\sigma = \sqrt{\frac{105}{50} - \left(\frac{29}{50}\right)^2} \times 10 = \sqrt{2.1 - 0.3364} \times 10 \approx 13.28$.

Modal class 20–30 ($f_1 = 15, f_0 = 8, f_2 = 13$): $M_o = 20 + \frac{15 - 8}{30 - 8 - 13} \times 10 = 20 + \frac{70}{9} \approx 27.78$.

c. $S_k = \frac{29.2 - 27.78}{13.28} \approx 0.11$: slightly positively skewed, nearly symmetric; a few higher earners pull the mean a little above the mode.

MARKING SCHEME

- ✓ a. Frequencies 4, 8, 15, 13, 6, 4.
- ✓ b. Mean 29.2.
- ✓ $\sigma \approx 13.28$.
- ✓ Mode about 27.78.
- ✓ c. $S_k \approx 0.11$.
- ✓ Slight positive skew, nearly symmetric.

4 When the two coefficients disagree

ANSWER HA · 6

Tests: the mode form rests on one clear peak with a tail; here the most common value sits at the top end, so the two forms (which only "usually" agree) disagree in sign. The tempting wrong answer: reporting whichever was computed first as "the" skew.

a. Mean $= \frac{42}{7} = 6$; median, the 4th value, **4**; mode **10**. $\Sigma x^2 = 338$, $\sigma = \sqrt{\frac{338}{7} - 36} = \sqrt{\frac{86}{7}} \approx 3.51$.

b. Mode form: $\frac{6 - 10}{3.51} \approx$ -1.14. Median form: $\frac{3(6 - 4)}{3.51} \approx 1.71$.

c. The data are two clusters, four small delays and three of 10 minutes; the "mode" 10 is not a peak with a tail on one side but the end of the set. Pearson's mode form assumes a single peak with a tail, so it misleads. The median form uses where the middle bus is: half the delays are 4 minutes or less, and the three long delays pull the mean above it, a tail to the right. **The median form, positive**, describes them better (the unit's rule: with an ill-defined mode, use the median).

MARKING SCHEME

- ✓ a. Mean 6, median 4, mode 10.
- ✓ $\sigma \approx 3.51$.
- ✓ b. Mode form about −1.14.
- ✓ Median form about 1.71.
- ✓ c. The mode is ill-defined here (two clusters, the "mode" at an end, not a peak with a tail).
- ✓ The median form is the better one: positively skewed.

5 Make your own

ANSWER

Tests: skew as where the tail is, built by hand: the mean above the median needs a few large values, and how large is fixed by the total. The tempting slips: thinking the sign depends on which set was written (mean and median alone fix it), and in b guessing from one example instead of using the total: the two smallest are at most 8 for the least largest value, and 0 for the greatest.

a. For example 6, 7, 8, 9, 20: total 50, so mean 10; median 8. For **every** such set $\overline{X} - M_d = 10 - 8 = 2 > 0$, and $\sigma > 0$ (the five aren't all equal), so $S_k = \frac{3(\overline{X} - M_d)}{\sigma} > 0$: **positively skewed**, whichever numbers were chosen.

b. Put them in order: $a \le b \le 8 \le d \le e$, with $a + b + d + e = 50 - 8 = 42$.

Science Playground, Katunje, Bhaktapur
Book a class visit: https://science-playground.com/booking

Least: a and b are at most 8, so $d + e \geq 42 - 16 = 26$; e is the larger of the two, so $e \geq 13$. It happens: **8, 8, 8, 13, 13**.

Greatest: a, $b \geq 0$ and $d \geq 8$, so $e \leq 42 - 8 = 34$. It happens: **0, 0, 8, 8, 34**. So the largest value is between **13 and 34**: a mean 2 above the median needs at least one value 5 above it.

c. Negative needs $\overline{X} < M_d$, so the median, the third number in order, is more than 10, and the fourth and fifth are at least as large: three numbers over 10, with small values making up the total of 50.

For example 2, 8, 13, 13, 14: mean 10, median 13, $S_k = \dfrac{3(10 - 13)}{\sigma} < 0$: **negatively skewed**, the 2 a tail to the left.

MARKING SCHEME

- ✓ a. A set with total 50 and median 8; positive for every set because $\overline{X} - M_d = 2 > 0$.
- ✓ b. $e \geq 13$ from $d + e \geq 26$.
- ✓ $e \leq 34$, with sets that reach both ends.
- ✓ c. A set with mean 10 and median over 10, and the reason three numbers exceed 10.

Marks by level	Knowledge	Understanding	Application	Higher ability	Total
Marks	0	0	12	16	**28**

Science Playground, Katunje, Bhaktapur
Book a class visit: https://science-playground.com/booking