CHARTA LEARNING · LESSON NOTES

Statistics and overlapping sets

Quantitative Reasoning

1 A story: two classes, one average

- Class 1: 70, 72, 74
- Class 2: 50, 72, 94
- Same average of 72, very different lists
- Middle and spread: two questions

IN THE LESSON

- Imagine two small classes that each took a test.
- In the first class the marks were seventy, seventy two and seventy four.
- In the second they were fifty, seventy two and ninety four.
- Both classes have the same average, seventy two.
- Yet you would not describe them the same way.
- In the first, everyone did about the same.
- In the second, one student struggled and one shone.
- So a single number is never the whole story.
- To describe a list properly we want to know where its middle is, and how far the values wander away from it.
- That is the first half of this lesson, and the exam loves testing it with very little calculation.

2 Mean, median, mode and range

- Mean: total divided by count
- Median: middle of the sorted list
- Mode: the most frequent value
- Range: largest minus smallest

IN THE LESSON

- Four words to define.
- The mean is the ordinary average.
- Add up all the values, then divide by how many there are.
- Think of tipping every pile of sweets onto a table and dealing them back out evenly.
- The median is the middle value once the list is sorted from smallest to largest.
- If the count is even, there are two middle values, so take the number halfway between them.
- The mode is the value that appears most often.
- A list can have one mode, several, or none.
- The range is the largest value minus the smallest, which tells you how wide the list stretches.
- Notice the median needs sorting.
- We will come back to that.

3 The rules for the middle

Mean = Total ÷ Count
Total = Mean × Count
Range = Largest − Smallest

Total of the list	Add every value
Count of values	How many values
Mean	The average

IN THE LESSON

- Here are the rules in their short form.
- The mean equals the total divided by the count.
- Let's take each term.
- The total is what you get when you add every value.
- The count is how many values there are.
- The mean is the answer.
- Now the useful twist.
- Turn the rule round and the total equals the mean times the count.
- That second form is the one the exam loves.
- If eight numbers average fifteen, their total is one hundred and twenty.
- Always ask what the total is, because the total is what changes when a number is added or taken away.
- The range needs no count: just the largest value take away the smallest.

4 Example 1: all four measures

Scores: 9, 7, 15, 10, 7, 12

Sorted list	7, 7, 9, 10, 12, 15	
Mean	60 ÷ 6	**10**
Median	(9 + 10) ÷ 2	**9.5**
Mode	7 appears twice	**7**
Range	15 − 7	**8**

IN THE LESSON

- Example one, a plain question.
- Six test scores are nine, seven, fifteen, ten, seven and twelve.
- Find the mean, median, mode and range.
- The first step is the one people skip: sort the list.
- Seven, seven, nine, ten, twelve, fifteen.
- Now the total.
- Nine plus seven plus fifteen plus ten plus seven plus twelve is sixty.
- Sixty divided by six is ten, so the mean is ten.
- There are six values, an even count, so the median is halfway between the third and fourth values, nine and ten, which is nine and a half.
- The mode is seven, because it appears twice.
- The range is fifteen minus seven, which is eight.

5 One big value pulls the mean

Value	
2	**2**
3	**3**
4	**4**
5	**5**
36	**36**

Total 50, so mean 10. The middle value is 4, so median 4.

IN THE LESSON

- Now a picture of why the mean and the median can disagree.
- Here are five values drawn as bars: two, three, four, five and thirty six.
- Along the bottom are the five values, and the height is the size of each.
- Four are small and one towers over the rest.
- The median is the middle bar once sorted, which is the third one, four.
- But the total is fifty, and fifty divided by five makes the mean ten.
- Not one of the five values is anywhere near ten!
- The big value dragged the mean upward, while the median stayed put.
- So when a question mentions an extreme value, remember the mean moves with it and the median hardly notices.

6 Example 2: working with totals

5 numbers average 14. Add a 6th: all 6 average 16

Old total	5 × 14	**70**
New total	6 × 16	**96**
The sixth number	96 − 70	**26**
Shortcut check	14 + 6 × 2	**26**

IN THE LESSON

- Example two, the kind of question that is typical on the exam.
- Five numbers have an average of fourteen.
- A sixth number is added, and the average of all six rises to sixteen.
- What is the sixth number?
- Do not hunt for the individual values.
- Think in totals.
- Five numbers averaging fourteen have a total of seventy.
- Six numbers averaging sixteen have a total of ninety six.
- The sixth number is whatever was added to the total, ninety six minus seventy, which is twenty six.
- Here is a faster check.
- The new number must lift each of the six numbers up by two, which costs twelve, and it must also supply the fourteen that it would have been on its own.
- Fourteen plus twelve is twenty six.

7 Example 3: the average of two averages

10 students average 70; 30 students average 90

First class total	10 × 70	**700**
Second class total	30 × 90	**2,700**
Everyone's total	700 + 2,700	**3,400**
Combined mean	3,400 ÷ 40	**85**
Tempting wrong answer	(70 + 90) ÷ 2	**80**

IN THE LESSON

- Example three is the trap version.
- One class of ten students averaged seventy.
- Another class of thirty students averaged ninety.
- What is the average of all forty students?
- The tempting answer is the middle of seventy and ninety, which is eighty.
- That would be right only if the classes were the same size.
- They are not, so the bigger class counts for more.
- Use totals again.
- The first class total is ten times seventy, seven hundred.
- The second is thirty times ninety, two thousand seven hundred.
- Together that is three thousand four hundred, spread over forty students, which is eighty five.
- A neat shortcut: the second class has three quarters of the students, so the answer sits three quarters of the way from seventy up to ninety.

8 Spread: how far from the middle?

- Spread: how far values sit from the mean
- Range uses only the two ends
- Standard deviation: the typical distance
- Judge distances, not the size of the numbers

IN THE LESSON

- Now the second question.
- How spread out is the list?
- Think of two archers who both aim for the bullseye.
- One puts all the arrows in a tight bunch.
- The other scatters them across the target.
- Their average position is the same, but they are very different archers.
- The simplest measure of spread is the range, but it only looks at two values.
- A better measure is the standard deviation.
- You do not need to calculate it on the exam.
- You need the idea.
- Standard deviation measures how far the values typically sit from the mean.
- Large distances from the mean give a large standard deviation, and small distances give a small one.
- So judge it by distances, not by the size of the numbers themselves.

9 Spread is about distances, not size

Spread (standard deviation)

40, 41, 42	**0.82**
140, 141, 142	**0.82**
1, 6, 11	**4.08**

Same gaps, same spread. Wider gaps, bigger spread.

IN THE LESSON

- Here is the idea in a picture.
- Three lists, and the height of each bar is its standard deviation.
- The first list is forty, forty one, forty two.
- The numbers are large, but they sit close together, each at most one away from the mean of forty one.
- Its bar is short, about point eight.
- The second list is one hundred and forty, one hundred and forty one, one hundred and forty two.
- Every number is a hundred bigger, but the gaps are identical, so the bar is exactly as short.
- The third list is one, six, eleven.
- The numbers are small, yet each sits up to five from its mean of six, so its bar is much taller, about four.
- Small numbers, big spread.

10 Example 4: which list is the most spread?

List 1: 3, 5, 7. List 2: 20, 22, 24. List 3: 1, 5, 9

List 1: mean 5	distances 2, 0, 2	**small**
List 2: mean 22	distances 2, 0, 2	**small**
List 3: mean 5	distances 4, 0, 4	**double**
Lists 1 and 2	same distances	**equal**
Greatest spread	the widest distances	**List 3**

IN THE LESSON

- Example four, a typical comparison, with no calculating at all.
- Which of three lists has the greatest standard deviation?
- List one is three, five, seven.
- List two is twenty, twenty two, twenty four.
- List three is one, five, nine.
- For each, find the mean and then the distance of each value from it.
- List one has a mean of five and distances of two, zero and two.
- List two has a mean of twenty two, and the distances are again two, zero and two.
- List three has a mean of five and distances of four, zero and four.
- Lists one and two have exactly the same spread, because list two is just list one shifted up.
- List three has distances twice as big, so it has the largest standard deviation, in fact exactly double.

11 Shift and scale rules

Add c: mean + c, spread same
Multiply by k: mean × k, spread × k

Add c to every value	Slide the whole list
Multiply by k	Stretch the whole list
New mean	Moves with both changes
New spread	Ignores shifts, scales by k

IN THE LESSON

- Here are the rules the exam tests most.
- Suppose you change every value in a list in the same way.
- First, add the same number c to every value.
- Every value moves along together, so the mean goes up by c, but the gaps between values do not change, so the spread stays exactly the same.
- Second, multiply every value by the same number k.
- The mean is multiplied by k, and the gaps are stretched by k as well, so the spread is multiplied by k, taken as a positive number.
- A negative k flips the list over, but a flipped list is just as spread out.
- The range obeys the same two rules as the standard deviation.

12 Example 5: change every value

Mean 20, standard deviation 4. New value = 3 × old − 5

New mean	3 × 20 − 5	**55**
New spread	3 × 4, ignore the −5	**12**
Wrong: 3 × 4 − 5	a shift never changes spread	**7**
A range of 10 becomes	3 × 10	**30**

IN THE LESSON

- Example five.
- A list of values has a mean of twenty and a standard deviation of four.
- Every value is now tripled and then reduced by five.
- What are the new mean and the new standard deviation?
- The mean follows both changes.
- Tripling gives sixty, and taking away five gives fifty five.
- The standard deviation only follows the stretching.
- Tripling makes it twelve.
- The reduction by five is a shift, and a shift does not touch the spread.
- So the new standard deviation is twelve, not seven.
- That wrong answer, three times four minus five, is exactly what the question writer hopes you will pick.
- If the question asked about the range of a list that started at ten, it would be thirty.

13 Example 6: adding one more value

Start with 6, 8, 10, 12, 14: mean 10, spread 2.83

Add 10, the mean itself	mean stays 10	**spread 2.58**
Add 40, far away	mean rises to 15	**spread 11.47**
Rule	near the mean: spread falls	
Rule	far from the mean: spread rises	

IN THE LESSON

- Example six asks what happens when one value joins the list.
- Take six, eight, ten, twelve and fourteen.
- The mean is ten, and the standard deviation is about two point eight three.
- First add another ten, right at the mean.
- The mean stays at ten.
- The new value is a distance of zero from the mean, so it makes the list more bunched, and the spread falls, to about two point five eight.
- Now instead add forty, far from the rest.
- The mean jumps to fifteen, and the spread leaps to about eleven and a half.
- The rule: a new value close to the mean shrinks the spread, and a new value far from the mean grows it.
- Often you just need that rule, not the numbers.

14 Mistakes to avoid with statistics

- 1. Sort the list before you take a median
- 2. Even count: halfway between the middle two
- 3. Different group sizes: use totals, not averages
- 4. Spread is about distances, not size
- 5. Adding a constant never changes the spread

IN THE LESSON

- Before the second half, five mistakes that cost marks.
- One.
- Finding a median without sorting.
- The middle of the list as written is not the median.
- Two.
- Forgetting that an even count has two middle values, so the median is halfway between them.
- Three.
- Averaging two averages when the groups are different sizes.
- Always go back to totals.
- Four.
- Judging spread by the size of the numbers.
- Look at the distances from the mean.
- Five.
- Thinking that adding the same number to every value changes the spread.
- It moves the mean and leaves the standard deviation alone.
- Keep these five in mind and you will pick up the easy marks.

15 A new story: counting people twice

- 60 in the club: 35 chess, 28 draughts
- 35 + 28 = 63, more than 60!
- People who play both are counted twice
- Remove the double count; remember those in neither

IN THE LESSON

- Now the second half of the lesson.
- Imagine a club of sixty people.
- Thirty five of them play chess, and twenty eight play draughts.
- Add thirty five and twenty eight and you get sixty three, which is more than the sixty people in the club!
- What went wrong?
- Some people play both games, and each of them has been counted twice, once in each group.
- That is the whole idea of overlapping sets.
- When groups share members, adding the group sizes counts the shared members too many times.
- So we have to take the double count off.
- We also need to remember the people who play neither game, who are in the club but in neither group.

16 The two-group rule

Total = A + B − Both + Neither

Group A	Everyone in the first group
Group B	Everyone in the second group
Both	Counted twice, so take off once
Neither	In the total, in neither group

IN THE LESSON

- Here is the rule for two overlapping groups.
- The total equals group A plus group B, minus both, plus neither.
- Let's take each piece.
- Group A is everyone in the first group, and group B is everyone in the second.
- Adding them counts the shared people twice, so we subtract both once, which is the number in both groups.
- That leaves everyone in at least one group, counted exactly once.
- Neither is the number in the total who belong to no group at all, and they are added at the end.
- Rearranged, this lets you find any one missing piece, which is what the exam usually asks.
- Fill in what you know, then solve for what you do not.

17 The Venn diagram: fill it in this order

1. The overlap first → 2. Each group only → 3. The neither region → 4. Add up: matches the total?

IN THE LESSON

- A Venn diagram is two overlapping circles inside a box.
- The left circle is group A, the right circle is group B, the overlap is both, and the box outside the circles is neither.
- The order you fill it in matters, and this is the method for every question.
- First, the overlap, because it is the region everything else depends on.
- Second, the regions that are in one group only, found by taking the overlap away from each group total.
- Third, the outside of the circles, the neither region.
- Fourth, add the four regions and check they come to the total.
- If the question does not give you the overlap, call it x and use the rule to solve for it.
- This order never lets you count anyone twice.

18 Example 7: the club, completed

60 people: 35 chess, 28 draughts, 12 both

Add the two groups	35 + 28	**63**
At least one game	63 − 12	**51**
Neither	60 − 51	**9**
Chess only	35 − 12	**23**
Draughts only	28 − 12	**16**

IN THE LESSON

- Example seven is the plain case.
- Sixty people, thirty five play chess, twenty eight play draughts, and twelve play both.
- How many play at least one game, and how many play neither?
- Add the two groups: thirty five plus twenty eight is sixty three.
- Take off the twelve counted twice, and you get fifty one people who play at least one game.
- So neither is sixty minus fifty one, which is nine.
- Now the diagram regions.
- Chess only is thirty five minus twelve, twenty three.
- Draughts only is twenty eight minus twelve, sixteen.
- The four regions are twenty three, twelve, sixteen and nine, and they add to sixty.
- The numbers all agree.

19 Example 8: find the overlap

120 customers: 70 tea, 65 coffee, 25 neither

At least one drink	120 − 25	**95**
Add the groups	70 + 65	**135**
Both	135 − 95	**40**
Tea only	70 − 40	**30**
Coffee only	65 − 40	**25**

IN THE LESSON

- Example eight is the typical version, where the overlap is the unknown.
- A shop has one hundred and twenty customers in a day.
- Seventy buy tea, sixty five buy coffee, and twenty five buy neither.
- How many buy both?
- Start with the people who bought something.
- Out of one hundred and twenty, twenty five bought neither, so ninety five bought at least one drink.
- Now add the two groups: seventy plus sixty five is one hundred and thirty five.
- That is forty more than ninety five, and the extra forty is exactly the people counted twice.
- So forty customers buy both.
- Check the regions.
- Tea only is thirty, coffee only is twenty five, both is forty, and neither is twenty five.
- They add to one hundred and twenty.

20 Example 8 as regions

Customers (120 in all)

Tea only	**30**
Both	**40**
Coffee only	**25**
Neither	**25**

Tea = 30 + 40 = 70. Coffee = 25 + 40 = 65.

IN THE LESSON

- Here are the four regions of the shop, as bars.
- Along the bottom are the four regions of the diagram, and the height is the number of customers in each.
- Tea only has thirty customers.
- Both has forty, the tallest bar, because the two groups overlap a lot.
- Coffee only has twenty five.
- Neither has twenty five.
- Notice that none of the regions overlaps another, so if you add the four bars you get the total exactly, one hundred and twenty.
- Notice also that the tea group is not just the tea only bar.
- The whole tea group is tea only plus both, thirty plus forty, which is seventy.
- Always remember a group total includes its overlap.

21 Three overlapping groups

- Exactly one group: counted once
- Exactly two groups: counted twice
- All three groups: counted three times
- Take off one for each pair, two for each triple

IN THE LESSON

- With three groups, the circles overlap in more ways, and the counting gets a little trickier.
- Picture three clubs: art, music and sport.
- Someone in exactly one club is counted once when you add the three group sizes.
- Someone in exactly two clubs is counted twice.
- Someone in all three is counted three times.
- We want every person counted exactly once.
- So the double counts have to come off.
- Each person in exactly two groups needs one subtraction.
- Each person in all three groups was counted three times, so they need two subtractions.
- That is where the two comes from in the rule, and it is where most mistakes happen.

22 The three-group rule

Total = A + B + C − Exactly two − 2 × All three + Neither

A + B + C	Sum of the three group sizes
Exactly two	In exactly two groups
All three	Taken off twice
Neither	Added at the end

IN THE LESSON

- Here is the rule for three groups.
- The total equals A plus B plus C, minus exactly two, minus two times all three, plus neither.
- Let's take each term.
- A plus B plus C is the sum of the three group sizes.
- Exactly two is the number of people in exactly two of the groups, not counting those in all three.
- All three is the number in every group, and it is taken off twice.
- Neither is added at the end.
- There is a second form of the same rule.
- When the question gives you the pair counts, where each pair count includes the people in all three, the rule reads: take off the pair counts, then add the all three number back once.

23 Example 9: three groups, exactly two

90 members: art 40, music 35, sport 30; exactly two 18; all three 5

Add the three groups	40 + 35 + 30	**105**
At least one	105 − 18 − 2 × 5	**77**
None of the three	90 − 77	**13**
Exactly one	77 − 18 − 5	**54**
Check by weight	54 + 2 × 18 + 3 × 5	**105**

IN THE LESSON

- Example nine.
- A society has ninety members.
- Forty join art, thirty five join music and thirty join sport.
- Eighteen belong to exactly two of the groups and five belong to all three.
- How many belong to none, and how many to exactly one?
- Add the three groups: forty plus thirty five plus thirty is one hundred and five.
- Take off eighteen once, and five twice.
- One hundred and five minus eighteen minus ten is seventy seven, the members in at least one group.
- So none is ninety minus seventy seven, thirteen.
- For exactly one, take the seventy seven and remove the eighteen and the five, leaving fifty four.
- Check it by weight.
- Fifty four once, eighteen twice and five three times is fifty four, plus thirty six, plus fifteen.
- That is one hundred and five.

24 Example 10: when pair counts are given

100 people: A 50, B 40, C 30. Pairs 15, 12, 10 (include all three). All three 6

Add the groups	50 + 40 + 30	**120**
Take off the pair counts	120 − 37	**83**
Add all three back once	83 + 6	**89**
Neither	100 − 89	**11**
Exactly two	37 − 3 × 6	**19**

IN THE LESSON

- Example ten is the trap.
- A hundred people.
- Fifty like A, forty like B and thirty like C.
- Fifteen like both A and B, twelve like both B and C, and ten like both A and C.
- Six like all three.
- Here the pair counts include the people who like all three, so we must not use the exactly two rule.
- Add the groups: one hundred and twenty.
- Take off the three pair counts, thirty seven, which leaves eighty three.
- But the six in all three were counted three times, then taken off three times, so they are now missing.
- Add them back once, eighty nine.
- Neither is eleven.
- And exactly two is thirty seven minus three times six, which is nineteen.
- Read the wording carefully.

25 Least and greatest overlap

Greatest overlap = smaller group
Least overlap = A + B − Total

Greatest overlap	The smaller group
Least overlap	Make neither as small as possible
Total	Everyone in the question

IN THE LESSON

- The exam also asks how large or how small an overlap could be.
- Here are the two limits.
- Take two groups inside a total.
- The overlap is as large as possible when the smaller group sits entirely inside the larger one.
- So the greatest overlap is the size of the smaller group.
- The overlap is as small as possible when as many people as possible are in at least one group, which means pushing the neither region down to zero.
- Then the two-group rule gives the least overlap: A plus B minus the total.
- If that comes out negative, the groups need not overlap at all, and the least overlap is zero.
- Notice the neither region only affects the least overlap.

26 Example 11: how large, how small?

80 people: 55 own a car, 48 own a bike

Greatest both	the smaller group	**48**
Least both, nobody in neither	55 + 48 − 80	**23**
Now 6 own neither	only 74 own something	
New least both	55 + 48 − 74	**29**
New greatest both	still the smaller group	**48**

IN THE LESSON

- Example eleven.
- Eighty people.
- Fifty five own a car and forty eight own a bike.
- What are the greatest and least numbers who could own both?
- For the greatest, the bike owners fit entirely inside the car owners, so it is forty eight, the smaller group.
- For the least, let nobody own neither.
- Fifty five plus forty eight is one hundred and three, which is twenty three more than eighty.
- So at least twenty three people must own both.
- Now the twist.
- Suppose six people are known to own neither.
- Then only seventy four people own something.
- The least overlap rises to fifty five plus forty eight minus seventy four, which is twenty nine.
- The greatest overlap is still forty eight.

27 Mistakes to avoid with overlaps

- 1. Remember the neither group
- 2. A group total includes its overlap
- 3. All three: taken off twice, not once
- 4. Exactly two is not the same as a pair count
- 5. Add the regions and check the total

IN THE LESSON

- Five more mistakes that cost marks.
- One.
- Forgetting the neither group.
- If the question gives a total, check whether some people belong to no group, and add them.
- Two.
- Reading a group total as the group only.
- The whole group of tea drinkers includes the ones who also drink coffee.
- Three.
- In a three group question, subtracting the all three number only once.
- Count everyone: exactly one, twice for exactly two, and three times for all three.
- Four.
- Mixing up exactly two with the pair counts.
- Read the wording to see whether the people in all three are already inside.
- Five.
- Always check by adding the regions to the total.

28 Recap: the five ideas

Statistics and sets in five rules

Centre	Total = mean × count
Spread	Distances from the mean
Two groups	A + B − Both + Neither
Three groups	− Exactly two − 2 × All three
Limits	Smaller group; neither 0

IN THE LESSON

- Let's pull it together.
- For the centre of a list, the mean is the total divided by the count, and the total is mean times count.
- Sort before you take the median.
- For spread, think distances from the mean.
- Adding to every value leaves the spread alone, and multiplying by k multiplies it by k.
- For two groups, the total is A plus B minus both plus neither, and the Venn order is overlap first, then group only, then neither.
- For three groups, take off exactly two once and all three twice.
- And for limits, the greatest overlap is the smaller group, and the least comes from making neither zero.
- With these five tools you can answer almost every question of this kind.

29 Try three quick questions

1. 5 numbers average 12; remove one, the rest average 10. Number removed?
   Answer: Total 60 − 40 = 20.
2. Larger spread: 5, 6, 7 or 50, 60, 70?
   Answer: 50, 60, 70: distances are 10 times bigger.
3. 40 people: 22 French, 18 Spanish, 8 neither. How many both?
   Answer: 22 + 18 + 8 − 40 = 8.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- Five numbers average twelve.
- When one is removed, the other four average ten.
- What was the number removed?
- The total was sixty, the new total is forty, so the number was twenty.
- Question two.
- Which list has the larger standard deviation: five, six, seven, or fifty, sixty, seventy?
- The second, because its distances from the mean are ten times bigger.
- Question three.
- Of forty people, twenty two speak French, eighteen speak Spanish, and eight speak neither.
- How many speak both?
- Twenty two plus eighteen plus eight is forty eight, which is eight more than forty.
- So eight people speak both.
- Well done.