15 QUANTITATIVE REASONING Statistics and Sets

Welcome back to quant! In this group of lessons, I'll focus on statistics, sets, combinatorics, tables and graphs, and geometry. I'll start in this lesson by discussing the basics of statistics and sets.

Statistics

For any given set of data, there are 6 computations, or determinations, that you are expected to be comfortable with and able to solve for: mean, median, mode, range, interquartile range, and probability. I'll define each of these briefly, then discuss some further considerations before I jump into practice questions.

The **mean** (or arithmetic mean) of a set of terms is the average of those terms. You can derive this average by adding up all terms and dividing the sum by the total number of terms.

The **median** of a set of terms is the number in the middle of the list created by ordering the terms from least to greatest. If the list has an even number of terms and therefore has no one "middle" term, the median is the average of the two terms in the middle.

The **mode** is the term that appears most frequently in the list. A given set may have no mode, one mode, or more than one mode.

The **range** is the positive difference between the greatest and least terms in the list.

The **interquartile range** is the positive difference between the 25th and 75th percentiles of the data in the set. I'll discuss quartiles and percentiles further below.

Finally, **probability** is the likelihood of choosing one or more specific terms out of the total group.

Let's take a closer look at each of these computations.

You're familiar with the usefulness of averages, or **means**. You also calculate them fairly routinely by adding up the terms in a set, then dividing by the number of terms. For example, the average of {10, 25, 55} is $\frac{90}{3}$, or 30. Note that the average of a set does NOT have to be a member of the set: 30 is the mean of the set, but 30 is not a member of the set.

Let's think about that calculation $\frac{90}{3} = 30$. What if you were asked what number you had to add to the set in order to make the average of the new set 50? You can use a variable: $\frac{10 + 25 + 55 + x}{4} = 50$. If you multiply both sides by 4 and add the known terms, you get $90 + x = 200$, so $x = 110$. But note that the key piece of information is that in order to generate a target average of 50 with 4 terms, the terms must add up to (4)(50), or 200.

That method is useful in situations where you are told that a set of 6 numbers has an average of 12. You don't want to set up an equation with 6 variables! Instead, you want to recognize that $\frac{\text{sum}}{6} = 12$, so the sum of the 6 terms must be 72.

Something conceptual to recognize about averages is that if you add to a set a new element that is equal to the current average of the set, the average of the new set will remain the same. For example, if a set of 5 numbers averages 20 (sum = 100), then adding the number 20 to the set will result in the new set of 6 numbers keeping the same average of 20 (sum = 120).

Finally, it's useful to recognize situations in which averages aren't as useful as other statistical measures. Averages can be highly skewed by outliers, numbers that are much less or much greater than other numbers in the set. For example, if you have 3 people with the ages of {8, 9, 13}, the average age of those people is 10. However, if a 70-year-old person enters the room, the average age jumps to 25, a number that doesn't do a very good job of representing the actual ages of the people in the room.

That's precisely why the **median** of a set can, at times, be more useful in describing a given set than its mean. The median of a set is less impacted by outliers. In the example of the ages of people in the room, the median of the original set {8, 9, 13} is 9, the middle value. In the set {8, 9, 13, 70}, the new median is the average of 9 and 13, or 11. The median of 11 does a better job of describing the ages of the people in the room than the mean of 25!

Note that the median of a set does NOT have to be a member of the set, as seen in the set {8, 9, 13, 70}, which has a median of 11.

Unlike the situation with averages, it is theoretically possible to determine the median of a set of values from knowing the value of just one member of the set! Consider a set of 35 terms arranged from least to greatest. Each "half" of the set is 17.5 numbers, which means you'll have two sets of 17 numbers on either side of the middle number—the 18th number from either end of the set—which is the median. If you happen to know that the value of the 18th term is, for example, 34, then the median of the set is 34!

Value of Term		34	
Term #	**1–17**	**18**	**19–35**

There's one other aspect of medians to be aware of. Let's say that you are presented with the following set of 5 terms ordered from least to greatest:

Value of Term	1		5		11
Term #	**1**	**2**	**3**	**4**	**5**

The median of the set is 5. But let's consider the range of possible values for the second and fourth terms in the set. The second term can be any of the following: {1, 2, 3, 4, 5}. Don't make the mistake of thinking that a number to the left of the median must be less than the median! It simply has to be less than or equal to the median. Similarly, the fourth term in the set must be greater than or equal to the median, so it can have any value from 5 through 11, inclusive. If a question asked for the least possible average of the 5-member set above, you would select the least possible values for the second and fourth terms to generate this set: {1, 1, 5, 5, 11}, which has a mean of $\frac{23}{5}$, or 4.6. If a question asked for the greatest possible average of the 5-member set above, you would select the greatest possible values for the second and fourth terms to generate this set: {1, 5, 5, 11, 11}, which has a mean of $\frac{33}{5}$, or 6.6.

The **mode** of a set is relatively easy to determine, though be aware that a set can have more than one mode. Unlike the case with the mean or median of a set, the mode of a set must be a member of the set. For example:

The set {1, 4, 4, 5, 7} has a mode of 4.
The set {–3, 5, 5, 9, 12, 12, 18} has modes of 5 and 12.

The **range** of a set is simply the greatest value minus the least value. The range of a set changes only if a new member added to the set is less than the current least term or greater than the current greatest term. For example:

The set {–4, 0, 9, 24} has a range of [24 – (–4)], or 28. That range changes only if you add to the set a term with a value of less than –4 or greater than 24.

There is a slightly more complicated measure of range called **interquartile range**, which is the positive difference between the 25th and 75th percentiles of the data in the set. Let's delve more deeply into the concepts of quartiles and percentiles.

As the name suggests, *quartiles* are quarters. Each quartile consists of one-fourth of the data points, starting with the lowest when the points are arranged in numerical order. Quartiles can also be described in percentile terms; if you think of a complete set of terms as 100% of the terms, then each quartile is composed of 25% of the terms in the set. The median of a set is the 50th percentile of the set.

For example, in a set of 20 terms arranged from least to greatest, terms 1–5 are the first quartile, terms 6–10 are the second quartile, terms 11–15 are the third quartile, and terms 16–20 are the fourth quartile. In this example, the "second highest value of the first quartile" would be the value of the fourth term. The "lowest value in the third quartile" would be the value of the 11th term.

For the purposes of the GRE, to generate the 25th percentile of a set of data, you find the median of the subset of numbers to the left of the set's median (50th percentile). To generate the 75th percentile of a set of data, you find the median of the subset of numbers to the right of the set's median (50th percentile). Let's look at two examples:

Example 1: What is the interquartile range of the set {3, 5, 5, 6, 9, 10, 14, 29}?

Step 1: find the median. This set is arranged from least to greatest, so the median of a set of 8 numbers is the average of the fourth and fifth numbers. In this case, the average of 6 and 9 is 7.5, so the median (50th percentile value) is 7.5.

List the numbers to the left of the median. Note that because the median is 7.5, you'd include 6 in the numbers to the left of the median: {3, 5, 5, 6}. The median of this set of 4 numbers is the average of the second and third numbers, so 5. That's the 25th percentile value.

List the numbers to the right of the median: {9, 10, 14, 29}. The median of this set is the average of 10 and 14, so 12. That's the 75th percentile value.

The interquartile range of this set is 12 – 5, or 7.

Example 2: What is the interquartile range of the set {–3, 0, 5, 6, 9, 15, 22}?

Step 1: the median in this case is the fourth number, so 6.
Step 2: the median of the numbers to the left of the median {–3, 0, 5} is 0, so 0 is the 25th percentile value.
Step 3: the median of the numbers to the right of the median {9, 15, 22} is 15, so 15 is the 75th percentile value.

The interquartile range of this set is 15 – 0, or 15.

The **probability** covered in this chapter is simple probability—we will look at advanced probability topics in the next chapter. Here are a few fundamental probability concepts to keep in mind.

Probability is expressed as a number from 0 through 1. A probability of 0 indicates an event that cannot happen, and a probability of 1 indicates an event that is certain to happen. Probabilities between 0 and 1 can be expressed as fractions (e.g., $\frac{1}{4}$), or as decimals (e.g., 0.25). Note that probabilities are not expressed using percentages—e.g., the probability of an event certain to happen is 1, not 100%.

The sum of all probabilities in a given situation must add up to 1. Example:

A bag contains only red, blue, and green marbles. If the probability of choosing a red marble is $\frac{1}{3}$ and the probability of choosing a blue marble is $\frac{1}{7}$, what is the probability of choosing a green marble?

$\frac{1}{3} + \frac{1}{7} + x = 1$

Use a common denominator: $\frac{7}{21} + \frac{3}{21} + x = \frac{21}{21}$

$x = \frac{11}{21}$, the probability of choosing a green marble

Note that this indicates that the minimum number of marbles in the bag is 21, and the number of marbles must be a multiple of 21. Otherwise, you'd have a non-integer number of green marbles, which is not possible.

The simplest way to calculate and express probability is to create a fraction in which the numerator is the number of elements or events that meet the conditions you're looking for ("working" elements or arrangements) and the denominator is the total number of elements or events possible. A shorthand way to express this fraction is $\frac{\text{working}}{\text{total}}$.

Let's practice with some sample questions.

High Temperatures

Below is a table of high temperatures in a certain area over the course of 2 full weeks:

Day	**Sun**	**Mon**	**Tues**	**Wed**	**Thur**	**Fri**	**Sat**	**Sun**	**Mon**	**Tues**	**Wed**	**Thur**	**Fri**	**Sat**
High Temp	78	80	82	84	83	82	76	72	70	68	71	72	78	82

Solve for the following for all 14 days:

Mean:
Median:
Mode:
Range:

Sample Questions

If 1 day during the 2 weeks is selected at random, what is the probability that the high for that day will be 80° or greater?

Ⓐ $\frac{1}{14}$

Ⓑ $\frac{1}{7}$

Ⓒ $\frac{2}{7}$

Ⓓ $\frac{5}{14}$

Ⓔ $\frac{3}{7}$

<u>Quantity A</u>	<u>Quantity B</u>
The average temperature of the 5 days with the lowest highs	The median temperature of the 5 days with the lowest highs

Ⓐ Quantity A is greater.
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

Dancing Competition

A dancer received the following scores from a set of 8 judges:

Judge:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Score:	9.0	9.4	8.8	9.4	8.9	9.5	9.4	9.2

Solve for the following:

Mean:
Median:
Mode:
Range:

Sample Questions

Quantity A	Quantity B
The range of the 5 lowest scores	The range of the 5 highest scores

Ⓐ Quantity A is greater.
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

If the competition has a 9th judge whose score is worth twice as much as that of each of the other 8 judges, what is the minimum score the dancer can receive from the 9th judge to maintain an average score of 9.0 or greater?

[ ]

High Temperatures Solutions

In Order:	68	70	71	72	72	76	78	78	80	82	82	82	83	84

Mean: 77

The average of all the terms is $\frac{68 + 70 + 71 + 72 + 72 + 76 + 78 + 78 + 80 + 82 + 82 + 82 + 83 + 84}{14} = 77$.

Median: 78

To find the median of 14 terms, take the average of the 7th and 8th greatest terms. In this case, both the 7th and 8th terms are 78, so 78 is the median.

Mode: 82

The mode is 82. 82 appears 3 times on the list, which is more times than any other term.

Range: 16

The highest high temperature was 84, and the lowest high temperature was 68. The difference between them is 84 – 68 = 16, which represents the range.

Sample Question Solutions

If 1 day during the 2 weeks is selected at random, what is the probability that the high for that day will be 80° or greater?

Ⓐ $\frac{1}{14}$

Ⓑ $\frac{1}{7}$

Ⓒ $\frac{2}{7}$

Ⓓ $\frac{5}{14}$

Ⓔ $\frac{3}{7}$

You can see that 6 of the 14 days had temperatures of 80° or greater. Here, $\frac{6}{14}$ can be reduced to $\frac{3}{7}$, and **(E) is the correct answer**.

Quantity A	Quantity B
The average temperature of the 5 days with the lowest highs	The median temperature of the 5 days with the lowest highs

Ⓐ Quantity A is greater.

Ⓑ Quantity B is greater.

Ⓒ The two quantities are equal.

Ⓓ The relationship cannot be determined from the information given.

The 5 lowest high temperatures are 68, 70, 71, 72, and 72. The median of these 5 temperatures is 71. You can infer that the average temperature must be lower than that because 68 and 70 are farther away from the median than are 72 and 72, but you can also calculate the average as follows: $\frac{68 + 70 + 71 + 72 + 72}{5} = 70.6$.

You can see that the median temperature is greater than the average temperature, so **Quantity B is greater**.

Dancing Competition Solutions

In Order:	8.8	8.9	9.0	9.2	9.4	9.4	9.4	9.5

Mean: 9.2
You can derive the average of the scores by adding them all up and dividing the sum by 8:

$$\frac{8.8 + 8.9 + 9.0 + 9.2 + 9.4 + 9.4 + 9.4 + 9.5}{8} = 9.2.$$

Median: 9.3
Here, because there are 8 scores on the list, the median is the average of the 4th and 5th scores, which are 9.2 and 9.4, respectively. The halfway point between those 2 scores is 9.3.

Mode: 9.4
The score 9.4 appears 3 times, which is more than any other score.

Range: 0.7
The lowest score is 8.8, and the highest is 9.5, so the range of scores is 9.5 – 8.8 = 0.7.

Sample Question Solutions

Quantity A	Quantity B
The range of the 5 lowest scores	The range of the 5 highest scores

Ⓐ **Quantity A is greater.**
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

For the 5 lowest scores, the range is 9.4 – 8.8 = 0.6.
For the 5 highest scores, the range is 9.5 – 9.2 = 0.3.
Therefore, **Quantity A is greater**.

If the competition has a 9th judge whose score is worth twice as much as that of each of the other 8 judges, what is the minimum score the dancer can receive from the 9th judge to maintain an average score of 9.0 or greater?

8.2

You can imagine the ninth judge's score as counting twice, which, added to the initial scores, would give you 10 scores in total.

For the 10 scores to average at least a 9.0, you need them to add up to at least 90.

The 8 scores you already have add up to 8.8 + 8.9 + 9.0 + 9.2 + 9.4 + 9.4 + 9.4 + 9.5 = 73.6.

To get to 90, you need the final 2 scores to be at least 90 – 73.6 = 16.4.

For those 2 scores to be 16.4, each must be at least $\frac{16.4}{2} = 8.2$

So, if the ninth judge gives at least an 8.2, their score will count double, which will add 16.4 to the total of all scores, and 16.4 + 73.6 = 90, which, divided by 10, gives you an average score of at least 9.0.

Again, **the correct answer is 8.2**.

Additional Considerations

Very commonly, you are asked to relate these stats to one another. You may also be asked to infer something about one stat or about the group as a whole from another stat. In doing so, there are two particular issues that require a second layer of thought:

Mean vs. Median

The GRE writers love creating questions that require you to consider the relationship between the mean and the median of a set of terms. To illustrate how to think about these questions, let's look at two different sample sets of terms, each presented in order:

List 1: 4, 6, 12, 13, 14

List 2: 11, 12, 12, 15, 19

Notice that, for both, the median is 12.

When you look at list 1, you can see that the numbers below the median are significantly farther away from it than are the numbers above the median. Thus, the average will be less than the median, and you can verify this with some computation:

$$\frac{4 + 6 + 12 + 13 + 14}{5} = 9.8$$

If you look at list 2, you can see that the numbers above the median are significantly farther away from it than are the numbers below the median. Thus, the average will be greater than the median, and you can again verify this with some computation:

$$\frac{11 + 12 + 12 + 15 + 19}{5} = 13.8$$

When asked to compare mean vs. median, you want to consider how numbers less than and greater than the median relate to it. If you notice a heavy bias one way or the other, it can tip off the relationship between the mean and the median.

Combining Averages

Sometimes problems will present you with the averages of two or more sets. It's very important to keep in mind that *you cannot simply average different averages*—the reason is that if the sets contain different numbers of items, then they need to be properly weighted.

Let's consider a situation that makes this concept clear: In a class that gives 10 equally weighted tests, a student scores an average of 70 on the first 8 tests, then scores an average of 90 on the final 2 tests. What is the student's overall test average?

You cannot simply average the two averages of 70 and 90 to generate a test score average of 80—that gives too much weight to the final 2 tests.

Instead, in situations with multiple averages, you always want to calculate the sum of all the test scores, then divide by the number of tests.

But how do you generate the sum of the scores if you don't know each individual score? Well, consider that the average formula is $\frac{\text{sum}}{\text{number of items}} = \text{average}$. In the case of the first 8 tests, you know that $\frac{\text{sum}}{8} = 70$, so if you multiply both sides by 8, you get a sum of 560 points. Similarly, the equation for the final 2 tests is $\frac{\text{sum}}{2} = 90$. So the overall sum of the student's scores is 560 + 160, or 740. Now, you can calculate the average across all 10 tests by doing $\frac{740}{10}$ = average, for an overall test average of 74. Note that the overall average of 74 is lower than the incorrect average of averages outcome of 80 because there were more tests in the lower-scoring category than in the higher, so the overall average was pulled down toward the average of the lower set.

Now, let's look at what, at first glance, appears to be a very different question, but it actually hinges on the same concept that *different averages cannot simply be averaged*.

Imagine that you drove for 1 hour at 20 miles per hour and then drove for 5 more hours at 100 miles per hour. If you wanted to compute the average speed for the entire trip, you can't just average 20 and 100 because you spent a lot more time driving at 100 miles per hour than you did at 20 miles per hour. You would have to come up with the average speed by first calculating the total distance traveled and then dividing it by the total number of hours.

Total distance traveled: 1(20) + 5(100) = 520 miles

Total number of hours: 1 + 5 = 6 hours

Therefore, the average speed would be $\frac{520}{6}$, which is close to 87 miles per hour.

GRE writers can be very clever in tempting you to improperly average averages, even against your better judgment. Here's a final illustration:

A man drives to work at 40 miles per hour. He returns by the same route at 60 miles per hour. What was his average speed for the trip?

Based on the limited information given, it might be very tempting to simply average 40 and 60 and say that his average speed was 50 miles per hour.

However, this would be incorrect.

And that's because even though it's true that he traveled the *same distance* both ways, it's not true that he spent the same amount of *time* traveling both ways. In fact, he must have spent more time traveling at 40 miles per hour than at 60 miles per hour because he was driving more slowly.

To illustrate with concrete terms, let's imagine that his work is 120 miles away (the math will work out the same regardless of the distance you choose).

In this case, he would travel a total of 240 miles going to work and coming back.

At 40 miles per hour, it would take him 3 hours to get to work. And at 60 miles per hour, it would take him 2 hours to get back home. So, he would spend 5 total hours traveling.

$$\frac{\text{total distance}}{\text{total time}} = \frac{240}{5} = 48 \text{ miles per hour}$$

So, his average speed for the trip would actually be 48 miles per hour.

Questions like these may leave you nervous that you have to look out for traps. You don't. Instead, what's most important is to remember that you should not average different averages. When confronted with the types of situations exemplified above, make sure to fully work out the math: always find the sum of all the terms to be averaged and divide by the total number of terms.

Practice Questions

1. A storage facility receives 2 types of packages: package type *A*, each weighing exactly x pounds, and package type *B*, each weighing exactly $x + 12$ pounds.

If the facility receives 10 of package type *A* and 20 of package type *B*, which of the following represents the average weight of all packages received that day?

Ⓐ $\frac{x + 6}{25}$

Ⓑ $x + 8$

Ⓒ $2x + 4$

Ⓓ $\frac{30x + 24}{30}$

Ⓔ $\frac{x}{2} + 8$

2. A list of terms includes the numbers x, x^2, and 0.75, and $0 < x < 1$. If the range of the terms is 0.5, what is the value of x?

x = [ ]

3. List *A*: $x - 4, x, x + 4, x + 8, x + 10$
List *B* (not shown) consists of 6 terms, all of which were derived by adding 4 to each term in list *A*.

Quantity A	Quantity B
The median of list *B*	The mean of list *B*

Ⓐ Quantity A is greater.
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

4. l, m, n, o, and p are all multiples of 3, and $0 < l < m < n < o < p$. The average of the 5 terms is 9.6.

Quantity A	Quantity B
p	18

Ⓐ Quantity A is greater.
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

5. A list has 4 numbers. The average of the 4 numbers is equal to the median and the range. If the greatest of the 4 numbers is 12 more than the average of the 4 numbers, what is the least of the 4 numbers?

[ ]

6. A certain experiment has 3 mutually exclusive potential outcomes. Their respective probabilities are x, y, and z. If $x = 2y$ and $y = 3z$, what is the value of y?

(A) $\frac{1}{10}$
(B) $\frac{1}{8}$
(C) $\frac{1}{5}$
(D) $\frac{1}{4}$
(E) $\frac{3}{10}$

7. List A: 22, 16, 12, 4, 18, 30

Which of the following sets of numbers has the same median as list A above?

Indicate <u>all</u> such sets.

[A] 41, –3, 0, –1, –5, 32
[B] 91, 36, 5, 9, 10, 24
[C] 16, 16, 16, 16, 18, 14
[D] 14, 20, 16, 17, 17, 15
[E] 26, 17, 16, 17, 19, 13

8. Mrs. Watson teaches 2 history classes. Her first class has a total of 23 students, whose average grade is 86. Her second class has a total of 26 students, whose average grade is 90.

Quantity A	Quantity B
The average grade for all 49 students	88

(A) Quantity A is greater.
(B) Quantity B is greater.
(C) The two quantities are equal.
(D) The relationship cannot be determined from the information given.

Practice Questions Continued

9. A certain online clothing store gives customers the option of rating items on a scale of 1 to 5 stars, with 1 star being the lowest rating and 5 being the highest rating. A certain item at the store received the following ratings from reviewers:

Number of Stars	Number of Reviewers
1	1
2	0
3	4
4	2
5	5

If the item were to get only 5-star reviews going forward, how many consecutive 5-star reviews would be required in order for the item to achieve an average star rating greater than 4?

Ⓐ 1
Ⓑ 2
Ⓒ 3
Ⓓ 4
Ⓔ 5

10. A list of terms has a range of 16 and includes the numbers 18, 23, and 29. Which of the following terms could also be on the list?

Indicate all such terms.

[A] 8
[B] 14
[C] 26
[D] 32
[E] 36

Practice Question Solutions

1. A storage facility receives 2 types of packages: package type *A*, each weighing exactly x pounds, and package type *B*, each weighing exactly $x + 12$ pounds.

If the facility receives 10 of package type *A* and 20 of package type *B*, which of the following represents the average weight of all packages received that day?

Ⓐ $\frac{x + 6}{25}$

Ⓑ $x + 8$

Ⓒ $2x + 4$

Ⓓ $\frac{30x + 24}{30}$

Ⓔ $\frac{x}{2} + 8$

Solution

To calculate the average weight of all the packages, you have to first come up with the total weight of all packages, and then divide that by the total number of packages.

The total weight of all packages should be equal to $10(x) + 20(x + 12)$, which $= 10x + 20x + 240 = 30x + 240$

And you will be dividing that by 30, since there are 30 packages in total.

$$\frac{30x + 240}{30} = x + 8$$

(B) is the correct answer.

2. A list of terms includes the numbers x, x^2, and 0.75, and $0 < x < 1$. If the range of the terms is 0.5, what is the value of x?

x = **.5**

Solution

As I've discussed in previous lessons, when numbers are between 0 and 1, they become lesser and lesser when squared and taken to greater and greater exponents. So, you know that x^2 must be less than x.

According to the given parameters, there are initially two ways to get to a range of 0.5: x could be the greatest term and x^2 the least, or 0.75 could be the greatest term and x^2 the least.

By trying out some numbers such as 0.8 or 0.9 (which, when squared, lead to 0.64 and 0.81, respectively), you can quickly infer that you can't get a range of 0.5 when x is the greatest term and x^2 the least. The only way for you to have a range of 0.5 is for 0.75 to be the greatest term and x^2 to be the least.

That means that, since the greatest term is 0.75 and the range is 0.5, you can use the following to solve for x^2, the least term:

$$x^2 = 0.75 - 0.5 = 0.25$$

And if $x^2 = 0.25$, $x = \sqrt{0.25}$ **= 0.5**

Solutions Continued

3. List A: $x - 4, x, x + 4, x + 8, x + 10$
List B (not shown) consists of 6 terms, all of which were derived by adding 4 to each term in list A.

Quantity A	Quantity B
The median of list B	The mean of list B

Ⓐ Quantity A is greater.
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

Solution

You can add 4 to each of the elements in list A and imagine list B as follows:

List B: $x, x + 4, x + 8, x + 12, x + 14$

For these 5 terms, the median is $x + 8$.

The average is $\frac{x + x + 4 + x + 8 + x + 12 + x + 14}{5} = \frac{5x + 38}{5} = x + \frac{38}{5}$.

You know that $\frac{38}{5}$ is less than 8, so the average must be less than the median.

Quantity A is greater.

Note that adding 4 to each element on the list did not change the relationship between the median and the mean. If you were able to see this upfront, you could have chosen to start your work by evaluating the initial terms given in list A.

4. l, m, n, o, and p are all multiples of 3, and $0 < l < m < n < o < p$. The average of the 5 terms is 9.6.

Quantity A	Quantity B
p	18

Ⓐ Quantity A is greater.
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

Solution

For the average of the 5 terms to equal 9.6, the 5 terms must add up to $(9.6)(5) = 48$.

Let's think of multiples of 3, in order: 3, 6, 9, 12, 15, 18, 21…

The numbers rise quickly, and you can see that you're fairly limited in terms of picking 5 terms that add to 48.

Trying out the first 5 terms, you can see that they add to $3 + 6 + 9 + 12 + 15 = 45$. The only way to get to a total of 48 is to keep the first 4 terms as is and to switch the last one to 18, so that the total list is: $3 + 6 + 9 + 12 + 18 = 48$. 5 terms that add to 48 give you an average of 9.6.

So, the greatest of the 5 terms must be 18, and the two quantities are therefore equal. **The correct answer is (C)**.

5. A list has 4 numbers. The average of the 4 numbers is equal to the median and the range. If the greatest of the 4 numbers is 12 more than the average of the 4 numbers, what is the least of the 4 numbers?

12

Solution

If the greatest of the 4 numbers is 12 more than the average and the average is equal to the median, then you know that the greatest of the numbers is 12 more than the median.

If the average is equal to the median, you can also infer that the 2 middle terms, when the numbers are placed in order, must be "equidistant" from the average, for if they weren't (e.g., if one term was 3 away from the average and the other 4 away), the average and the median wouldn't equal one another.

If the 2 middle terms must be equidistant from the average, then so must the 2 outer terms—the least and the greatest (if the other 2 terms weren't equidistant from the average, you'd then arrive at a different average).

Since the greatest term is 12 greater than the average, the least must also be 12 less than the average. So, the range must be 24, and, per the initial conditions, you know that the mean and median must also be 24.

If the least term is 12 less than 24, it must equal 24 – 12 = 12.

6. A certain experiment has 3 mutually exclusive potential outcomes. Their respective probabilities are x, y, and z. If $x = 2y$ and $y = 3z$, what is the value of y?

Ⓐ $\frac{1}{10}$

Ⓑ $\frac{1}{8}$

Ⓒ $\frac{1}{5}$

Ⓓ $\frac{1}{4}$

Ⓔ $\mathbf{\frac{3}{10}}$

Solution

You know that all possible outcome probabilities must add up to 100%, or fractionally speaking, $\frac{1}{1}$.

So, $x + y + z = 100\%$, or $x + y + z = 1$ (whichever you prefer).

If $x = 2y$ and $y = 3z$, you know that x must equal $6z(x = 2(3z))$.

So, $x + y + z = 6z + 3z + z = 100\%$; $10z = 100\%$; $z = 10\%$.

If $z = 10\%$, $y = 30\%$ and $x = 60\%$.

Since you are asked to solve for y, and 30% is equivalent to $\frac{3}{10}$, **(E) is the correct answer**.

Solutions Continued

7. List A: 22, 16, 12, 4, 18, 30

Which of the following sets of numbers has the same median as list A above?

Indicate all such sets.

[A] 41, –3, 0, –1, –5, 32
[B] 91, 36, 5, 9, 10, 24
[C] 16, 16, 16, 16, 18, 14
[D] 14, 20, 16, 17, 17, 15
[E] 26, 17, 16, 17, 19, 13

Solution

When you put the elements in list A in order, 22, 16, 12, 4, 18, 30 becomes 4, 12, 16, 18, 22, 30, then you can see that the median = $\frac{16+18}{2} = 17$.

Let's put each of the answer choices in order to see what their respective medians are:

[A] 41, –3, 0, –1, –5, 32 becomes –5, –3, –1, 0, 32, 41, which has a median of $\frac{-1+0}{2} = -0.5$.
[B] 91, 36, 5, 9, 10, 24 becomes 5, 9, 10, 24, 36, 91, which has a median of $\frac{10+24}{2} = 17$.
[C]16, 16, 16, 16, 18, 14 becomes 14, 16, 16, 16, 16, 18, which has a median of 16.
[D] 14, 20, 16, 17, 17, 15 becomes 14, 15, 16, 17, 17, 20, which has a median of $\frac{16+17}{2} = 16.5$.
[E] 26, 17, 16, 17, 19, 13 becomes 13, 16, 17, 17, 19, 26, which has a median of 17.

Therefore, answer choices [B] and [E] have medians that equal the median of list *A* and are the correct answers.

8. Mrs. Watson teaches 2 history classes. Her first class has a total of 23 students, whose average grade is 86. Her second class has a total of 26 students, whose average grade is 90.

Quantity A	Quantity B
The average grade for all 49 students	88

(A) Quantity A is greater.
(B) Quantity B is greater.
(C) The two quantities are equal.
(D) The relationship cannot be determined from the information given.

Solution

Since 88 is halfway between 86 and 90, and you have more students at 90 than at 86, you can infer that the average will be greater than 88 and that Quantity A will therefore be greater.

If you are not sure, or would like to verify, you could calculate the average as follows:

$$\frac{(23)(86)+(26)(90)}{49} \approx 88.12$$

Again, Quantity A is greater.

9. A certain online clothing store gives customers the option of rating items on a scale of 1 to 5 stars, with 1 star being the lowest rating and 5 being the highest rating. A certain item at the store received the following ratings from reviewers:

Number of Stars	Number of Reviewers
1	1
2	0
3	4
4	2
5	5

If the item were to get only 5-star reviews going forward, how many consecutive 5-star reviews would be required in order for the item to achieve an average star rating greater than 4?

Ⓐ 1
Ⓑ 2
Ⓒ 3
Ⓓ 4
Ⓔ 5

<u>Solution</u>

Let's start by computing the current total number of stars, the number of ratings, and the average.

There are (1)(1) + (0)(2) + (4)(3) + (2)(4) + (5)(5) = 46 = 46 total stars.

There are 1 + 0 + 4 + 2 + 5 = 12 total ratings.

The current average is $\frac{46}{12}$, which is about 3.8.

If the store receives another 5-star rating, the average would go up to $\frac{51}{13}$, which is about 3.9.

If the store receives 2 more 5-star ratings, the average would go up to $\frac{56}{14}$, which is exactly 4.0.

Since you need an average higher than 4.0, you know you need more than 2 additional 5-star ratings, **so (C) is the correct answer**.

With three 5-star ratings, the average would go up to $\frac{61}{15}$, which is a little more than 4.

You can also set up an inequality to get the answer. Piggy-backing off some of the work you've done so far, knowing that you have 14 reviews that add up to 56 stars, that all the new reviews would be 5 stars, and you need an average greater than 4, you can set up the following inequality:

$$\frac{46 + 5x}{12 + x} > 4$$

This tells you to look for a number of 5-star reviews, which, when added to what you already have, will give you an average greater than 4.

You can solve the equation for x:

$$\frac{46 + 5x}{12 + x} > 4$$
$$46 + 5x > 4(12 + x)$$
$$46 + 5x > 48 + 4x$$
$$x > 2$$

This also gives you the correct answer of 3, which is (C).

Solutions Continued

10. A list of terms has a range of 16 and includes the numbers 18, 23, and 29. Which of the following terms could also be on the list?

Indicate all such terms.

[A] 8
[B] 14
[C] 26
[D] 32
[E] 36

Solution

It can often be helpful to take a moment to think about extremes—the least or greatest values based upon the information given to us, and that's true here.

If the list has a range of 16 and you imagine that 18 is the least number on the list, the greatest possible would be 18 + 16 = 34.

If you imagine that 29 is the greatest number on the list, the least possible number would be 29 – 16 = 13.

Taken together, you know that other numbers on the list must be greater than or equal to 13 and less than or equal to 34.

Answer choices [B], [C], and [D] fit into that range and are therefore correct.

Sets

Now, let's talk briefly about sets, or, more specifically, the issue of overlapping sets.

A set is simply a group of terms. If a question mentions two or more sets, it is possible that the sets are **mutually exclusive**, in which case no term can be included in more than one set. I'll say more about this situation when I discuss advanced probability in the next chapter.

Sometimes, a term may be a member of more than one set. Here are some indications that you are dealing with overlapping sets:

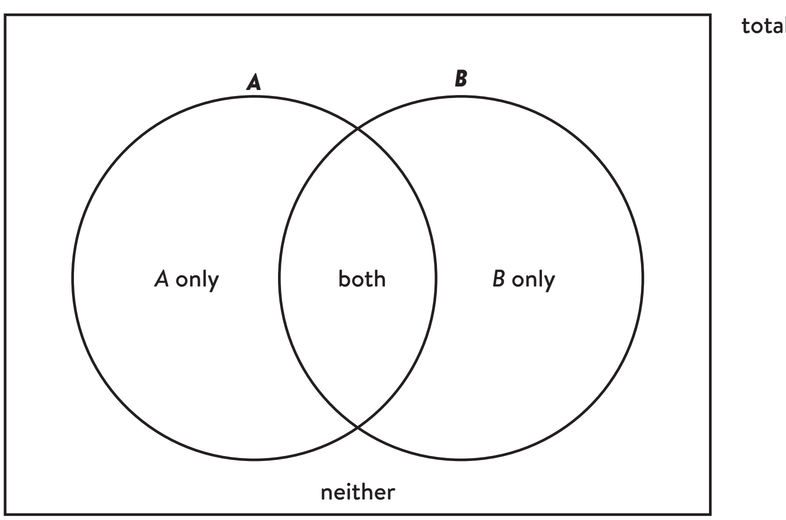
The sum of the set totals is greater than the total number of terms or elements in the problem. For example:

In a group of 50 students, 30 take math and 40 take physics. Well, 30 + 40 = 70, and there are only 50 students, so some students must be taking both (and are therefore included in both set totals).

The question uses the words *both* and/or *neither*, as in "How many students take *both* classes?" or "How many students take *neither* class?"

When sets potentially overlap, it is helpful to represent the situation with a **Venn diagram**, in which overlapping circles indicate the possibility that terms might be members of only one set, both sets, or neither set. Note that in an earlier chapter you used Venn diagrams to help you find the **greatest common factor** and **least common multiple** for two numbers; in this chapter, you'll use them to solve problems asking you to identify how many elements are in each possible position when sets overlap.

Here's a basic Venn diagram showing the potential overlap of two sets labeled *A* and *B*.



The diagram helps remind you that the total number or terms or elements in set *A*—which should be written *above* or *below* the circle—does not indicate whether those elements are in set *A* only or in both set *A* and set *B*. It also reminds you that it is possible for some elements to be in neither set.

The basic Venn diagram formula, using the notation from the diagram above, is **_A_ + _B_ – both + neither = total**

Let's look more closely at this formula, which efficiently solves many problems involving overlapping sets. Remember that the value of *A* (the total number of elements in circle *A*) can be expressed as "*A* only + both," and the value of *B* (the total number of elements in circle *B*) can be expressed as "*B* only + both." So *A* + *B* can be expressed as "*A* only + both + *B* only + both," which *double counts the elements in both sets*, and that's why the basic Venn formula subtracts the "both" quantity *once*, to eliminate that double counting. Once you add in the neithers, which are not included anywhere in the *A* or *B* circles, then you have the total number of elements.

Note that if you are asked, "How many elements are in sets *A* or *B*?," that "or" is *inclusive*, meaning that the elements in both sets should be included in your answer. Every element except the neithers is in set *A* or *B*. The phrase "*A* or *B*" is the same as the phrase "*A* or *B* or both" (which the test writers can also use).

The calculation *A* + *B* – both will give you the number of elements in *A* or *B* or both. In formal notation, $A \cup B$ (the union of sets *A* and *B*) is equivalent to "*A* or *B* or both."

If the test writers want to *exclude* the elements in both sets, in other words, to find the sum "*A* only + *B* only," they will ask a question phrased in one of the following ways (not a complete list):

"How many elements are in sets *A* or *B*, but not both?"
"How many elements are in exactly one set?"

Using Venn diagram notation, the formula for finding *A* or *B*, but not both, would be:

A + *B* – 2(both) = *A* only + *B* only

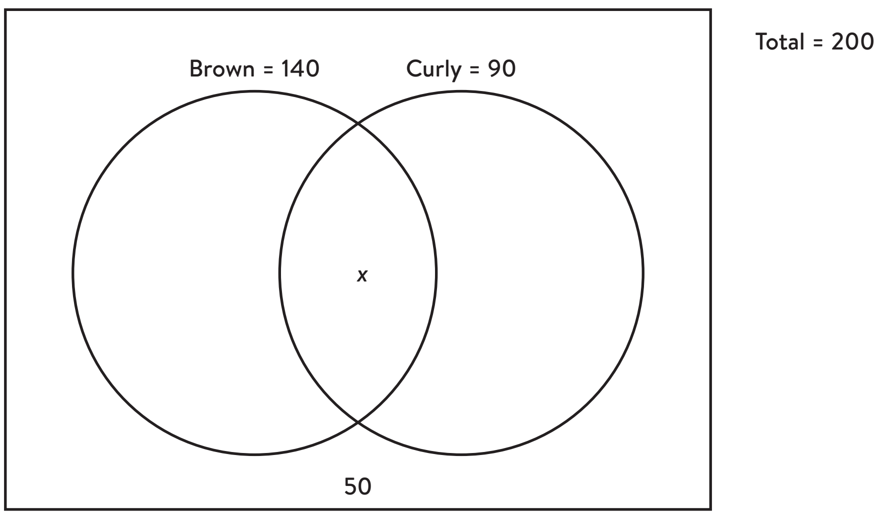
Remember that "$A + B$" double counts the boths, so if you want to remove them entirely, you have to subtract them *twice* from the sum $A + B$.

The formal notation $A \cup B$ (the intersection of sets A and B) is the same as the "both" category on the basic diagram.

Let's look at the context for a typical problem involving overlapping sets.

There are 200 people in a room. 140 of them have brown hair, 90 have curly hair, and 50 have neither brown nor curly hair.

140 + 90 = 230, and there are only 200 people in total, so the brown hair and curly hair sets must overlap. The word "neither" is also an indicator of a problem in which a Venn diagram can be useful.

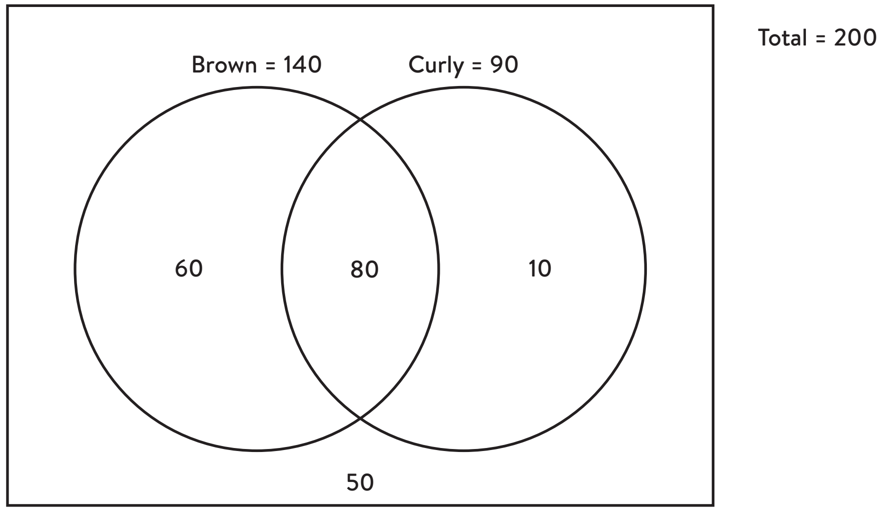


Let's use the formula $A + B -$ both + neither = total

$140 + 90 - x + 50 = 200$

$280 - x = 200$, so $x = 80$. 80 people have both brown and curly hair.

Depending on the question asked, it might be helpful to fill in all of the positions in your Venn diagram:



Note that once all of the non-overlapping positions are filled in, you can check your work by making sure that A only + both + B only + neither = total. In this case, 60 + 80 + 10 + 50 = 200, so that's good.

You can now answer any question asked, including "How many people have brown hair or curly hair, but not both?" That would be A only + B only, or 60 + 10 = 70, the same answer you'd get if you did $A + B - 2(\text{both})$, 140 + 90 – 2(80) = 70.

Overlapping set questions are often presented entirely in percentage terms. For example: Of the students in a school, 75% have brown hair, 40% have curly hair, and 5% have neither brown nor curly hair. What percent of the students have both brown and curly hair? The best way to handle these questions is to set the total number of students at 100, which turns 75% into the number 75, etc. So A is 75, B is 40, and neither is 5. 75 + 40 – both + 5 = 100, so both = 20, which you then convert back to 20%.

Sometimes the amount of overlap between two sets is not a fixed quantity but rather something that the test writers will ask you to maximize or minimize. A good indicator of this type of problem is phrasing such as "What is the greatest (or least) number of people who could be in both sets (or neither set)?"

In addressing such problems, let's start with some theoretical basics. You'll note below that the categories of both and neither are linked in a way that might initially seem strange: when you increase the boths, you increase the neithers (and vice versa), and when you decrease the boths, you decrease the neithers (and vice versa).

Maximizing the Boths and Neithers

The greatest number of elements that can be in both sets is the lesser of the two circle totals A and B. For example, if in a group of 100 students, 30 take math and 60 take physics, then the maximum number of students who could be in both classes is 30, the lesser of the two circle totals. In that case, you have 0 math only, 30 both, 30 physics only, and 40 neithers. Note that maximizing the boths also maximized the neithers.

Minimizing the Boths and Neithers

The number of elements in both sets can be 0 (mutually exclusive sets), but only if the sum of the two set totals (the circle totals we've been calling A and B) is less than or equal to the overall total. For example, if in a group of 100 students, 30 take math and 60 take physics, it's possible that there is 0 overlap between the groups: 30 take math only, 60 take physics only, and 10 take neither.

If the sum of the two set totals is greater than the overall total, then it is not possible for the boths to be 0. In this case, the way to minimize the boths is to set the neithers to 0, then use the formula to solve for the boths. For example, if in a group of 100 students, 50 take math and 70 take physics, then the boths can't be 0. So, set the neithers to 0 and use the formula: 50 + 70 – both + 0 = 100.

120 – both = 100

Both = 20. This is the minimum possible value for the boths.

Let's look at a sample problem:

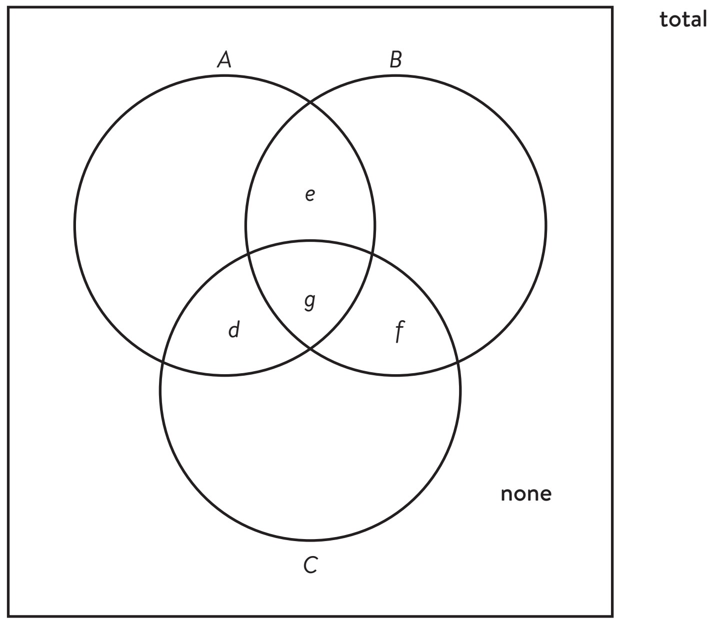
In a certain school, the Math Club has 40 members, and the Physics Club has 47 members. At least 3 members of the Math Club are not in the Physics Club. Which of the following could be the number of members in the Physics Club who are not in the Math Club?

Let's start by setting the value of the Math Club-only students at the minimum of 3. The other 37 members of the Math Club can be assigned to both clubs (because the number 37 is less than the Physics Club set total of 47), leaving 10 students in Physics Club only.

Now, let's maximize the Math Club-only students at 40, leaving 0 students in both clubs. All 47 Physics club students are therefore in Physics Club only.

The range of possible values for Physics Club-only students (the ones in Physics club but not in Math Club) is from 10 through 47. The format of this question would probably be multiple choice (all that apply), and you would choose all answers in that range.

It's possible, though unlikely, that a GRE question will involve 3 potentially overlapping sets. In that case, use the following Venn diagram:



The formula for this diagram is $A + B + C - (d + e + f) - 2(g) + \text{none} = \text{total}$

This formula operates on principles similar to the 2-circle formula. $A + B + C$ double counts the elements in positions d, e, and f (2-circle overlaps) and triple counts the elements in position g (3-circle overlaps). Therefore, you need to subtract the double counted elements once and the triple-counted elements twice to make sure that everything is counted only once.

Note that if you are told that 40 elements are in both sets A and B, those 40 elements could (unless otherwise noted) be in positions e (the overlap of just A and B) or g (the overlap of all 3 sets). So, if you know that there are 5 elements in position g, subtract those 5 from 40 to generate a total of 35 items in position e.

Practice Questions

1. 18 people attended a cookout. Each person ate a burger, a hot dog, or both. If 12 people ate a burger and 12 people ate a hot dog, how many people ate both burgers and hot dogs?

Ⓐ 4
Ⓑ 6
Ⓒ 8
Ⓓ 10
Ⓔ It cannot be determined.

2. Set *A*: All multiples of 4 less than or equal to 1,000
Set *B*: All multiples of 5 less than or equal to 1,000

Quantity A	Quantity B
The number of terms that are in set *A* that are not in set *B*	The number of terms that are in set *B* that are not in set *A*

Ⓐ Quantity A is greater.
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

3. There are 623 people in the Fairfax County database who have received a driving ticket, a parking ticket, or both in the past year. If at least 400 of those people received a parking ticket and an equal or greater number of them received a driving ticket, which of the following could be the number of drivers who received both a driving ticket and a parking ticket in the past year?

Indicate <u>all</u> such values.

[A] 136
[B] 187
[C] 223
[D] 426
[E] 623

Practice Question Solutions

1. 18 people attended a cookout. Each person ate a burger, a hot dog, or both. If 12 people ate a burger and 12 people ate a hot dog, how many people ate both burgers and hot dogs?

Ⓐ 4
Ⓑ 6
Ⓒ 8
Ⓓ 10
Ⓔ It cannot be determined.

Solution

If you add up the number of burgers and hot dogs eaten, you get 12 + 12 = 24, which is 6 more than the 18 people who attended. This must mean that these 6 people must have eaten both a burger and a hot dog, and **(B) is the correct answer**.

---

2. Set A: All multiples of 4 less than or equal to 1,000
Set B: All multiples of 5 less than or equal to 1,000

Quantity A	Quantity B
The number of terms that are in set A that are not in set B	The number of terms that are in set B that are not in set A

Ⓐ Quantity A is greater.
Ⓑ Quantity B is greater.
Ⓒ The two quantities are equal.
Ⓓ The relationship cannot be determined from the information given.

Solution

If you divide 1,000 by 4, you get 250. This tells you that there are 250 multiples of 4 that are less than or equal to 1,000.

If you divide 1,000 by 5, you get 200. This tells you that there are 200 multiples of 5 that are less than or equal to 1,000.

The terms that are elements in both lists are those that are multiples of both 4 and 5: multiples of 20.

If you divide 1,000 by 20, you get 50. This tells you that there are 50 multiples of 20 that are less than or equal to 1,000.

The number of terms that are in set A that are not in set B is therefore 250 – 50 = 200.

Similarly, the number of terms that are in set B that are not in set A is therefore 200 – 50 = 150.

The correct answer is (A).

---

3. There are 623 people in the Fairfax County database who have received a driving ticket, a parking ticket, or both in the past year. If at least 400 of those people received a parking ticket and an equal or greater number of them received a driving ticket, which of the following could be the number of drivers who received both a driving ticket and a parking ticket in the past year?

Indicate all such values.

[A] 136
[B] 187
[C] 223
[D] 426
[E] 623

Solution

If at least 400 people also received a driving ticket, you know that the overlap between the two categories must be at least 400 + 400 – 623 = 177.

In terms of the maximum overlap, there is nothing in the question stem that makes it impossible for all 623 people in the database to have received both a driving and parking ticket.

Therefore, answer choices [B], [C], [D], and [E] could all be viable numbers for the overlap.