# Inclusion/Exclusion Principle

Graham Middle School Math Olympiad Team

## BASIC STATISTICAL TERMS

The **average** (**arithmetic mean**) of a set of $N$ numbers is the sum of all $N$ numbers divided by $N$.

The **mode** of a set of numbers is the number occurring *most frequently* in the set.

The **median** of an ordered set (from smallest to largest) of $N$ numbers is the *middle number* if $N$ is odd, or the *mean of the two middle numbers* if $N$ is even.

The **range** of a set is a difference between the largest and the smallest elements.
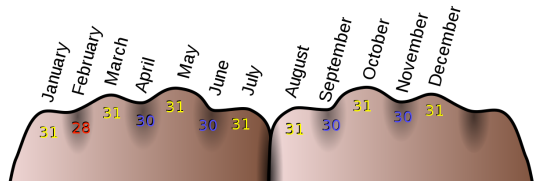


Consider the set containing the number of days in each month of the year (non-leap year). What are the mean, mode, and median for this set? The elements in this set are

$$\{31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31\}.$$

To find mode and median it is useful to sort them

$$\{28, 30, 30, 30, 30, 31, 31, 31, 31, 31, 31, 31\}.$$

31 occurs most frequently, so it is the **mode**.
The median is the mean of the 6th and 7th elements of the set (both of which are 31), so the **median** is 31.
The **mean** is the sum of all numbers in the set 365 divided by 12, which is $30\frac{5}{12}$.

The range is difference between 31 (the largest value) and 28 (the smallest), so the **range** is 3.

In mathematics, a **set** is a *collection* of elements.

The elements that make up a set can be any kind of mathematical objects: numbers, symbols, points in space, lines, etc. Sets are usually denoted as capital Latin letters.

Sometimes it is helpful to visualize sets as circles with all their elements inside. For example, this diagram represents 3 sets: the red set with elements 2, 5, and 1, the yellow set with elements 1 and 6, and the green with elements 4 and 7. 3 doesn't belong to any set. If we denote the sets as $R$, $Y$ and $G$, we can write

$$R = \{1,\ 2,\ 5\}, \quad Y = \{1,\ 6\} \quad \text{and} \quad G = \{4,\ 7\}.$$

The **number** of elements (the set *cardinality*) in a set $S$ is denoted as $|S|$.

$$|R| = 3, \quad |Y| = 2, \quad |G| = 2.$$
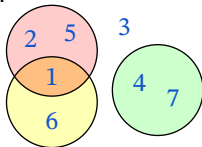
The **union** (denoted by $\cup$) of sets is the set of all elements in *any* of this set.

For example

$$R \cup Y = \{1,\ 2,\ 5,\ 6\}, \quad Y \cup G = \{1,\ 4,\ 6,\ 7\}.$$

The **intersection** (denoted by $\cap$) of sets is the set whose elements are in *all* sets.

For example

$$R \cap Y = \{1\}, \quad Y \cap G = \{\ \} = \varnothing.$$

The symbol $\varnothing$ denotes the **empty set**, the set without elements.

The **complement** of a set $S$ (denoted as $S'$ or $\overline{S}$) is a set of all elements *not in* a set $S$.

For example if we consider only elements from 1 to 7 (it is also called an **universal set**)

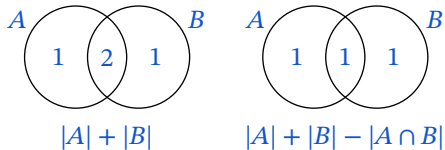$$R' = \{3,\ 4,\ 6,\ 7\}, \quad G' = \{1,\ 2,\ 3,\ 5,\ 6\}.$$

What the number of integers less than or equal to 100 that were factors of 4 or factors of 5?

We knew there were 25 factors of 4 and 20 factors of 5, but the final answer was

$$25 + 20 - 5 = 40,$$

because we had to subtract the 5 factors of 20 that were being counted twice.

We can generalize this to the counting of elements in any two sets $A$ and $B$.



$$|A| + |B| \qquad |A| + |B| - |A \cap B|$$

The numbers in the diagram above indicate how many times the elements in each region are counted.

Where $A$ and $B$ overlap, elements are counted twice.

Therefore to find the number of the elements in $A$ and $B$, we must subtract the number of the elements present in both sets.
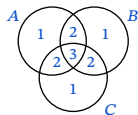
$$|A \cup B| = |A| + |B| - |A \cap B|.$$

There are 12 cookies with chocolate chips, and 10 cookies with nuts. If there are 4 cookies with chips and nuts, how many cookies are there?

To find the total number of cookies, we sum the cookies from each category ($12 + 10$), and then subtract the cookies that overlap both categories ($4$) to find the total is 18.

## INCLUSION/EXCLUSION PRINCIPLE EXTENDED

This counting principle for overlapping sets can be extended to 3 sets.

The diagram below shows 3 overlapping sets and the numbers labeling each region indicate how many times that region is counted in the operation shown beneath it. To find the total number of elements, we must count the elements in each set, then subtract the elements for the 3 pairwise intersections, and finally add back the elements found in all 3 sets.

$|A| + |B| + |C|$
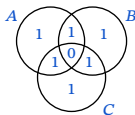
$|A| + |B| + |C| -$
$- |A \cap B| - |A \cap C| -$
$- |B \cap C|$
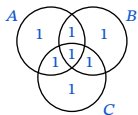
$|A| + |B| + |C| -$
$- |A \cap B| - |A \cap C| -$
$- |B \cap C| + |A \cap B \cap C|$

Again, the numbers in the diagrams above indicate how many times the elements in each region are counted.

There are 12 cookies with chocolate chips, 18 cookies with nuts, and 8 cookies with raisins. If there are 8 cookies with chips and nuts, 2 cookies with chips and raisins, 8 cookies with raisins and nuts, and 2 cookies with chips, nuts, and raisins, how many cookies are there in total?

Using the overlapping counting principle:

$$\text{chips} + \text{nuts} + \text{raisins} = 38,$$
$$\text{chips/nuts} + \text{chips/raisins} + \text{nuts/raisins} = 18,$$
$$\text{chips/nuts/raisins} = 2,$$
$$\text{Total number of cookies} = 38 - 18 + 2 = 22.$$

Casework is a counting method where you divide a problem into *several groups* based on some property (known as cases), count these cases individually, and then sum the totals from each case. Casework is a very general problem solving approach, and is used a tremendous amount in middle school math contests. Be careful of potential case overlap.
*If the same solution occurs in two of your cases, don't count it twice.*

How many pairs of positive integers $(x, y)$ satisfy the equation $x^4 + y < 70$?

**Case 1**: $x = 1$. $1 + y < 70$. So $0 < y < 69$, and there are $68$ possible pairs $(1, 1)$, $(1, 2)$, ..., $(1, 68)$.
**Case 2**: $x = 2$. $16 + y < 70$. So $0 < y < 54$, and there are $53$ possible pairs $(2, 1)$, $(2, 2)$, ..., $(2, 53)$.
Note $3^4 = 81$, so there are only two cases to consider ($x = 1$ or $2$) with no overlap between them.
$$\text{Total pairs} = 68 + 53 = 121.$$

There are a $3$ distinct urns (containers $A$, $B$, $C$) and $4$ identical balls. I need the put all of the balls into the urns, and I am free to place $0 - 4$ balls into any urn. How many distinct ways are there to put the balls into the urns?

**Case 1**: All $4$ balls in a single urn – $3$ **ways**.
**Case 2**: $3$ in one urn, $1$ in a different urn. There are $3$ choices for the urn with $3$ balls, and then $2$ choices remaining for the urn with the single ball, so there are $3 \times 2 = 6$ **ways**.
**Case 3**: $2$ balls in one urn, and $2$ balls in a different urn. It is $6$ arrangements ($3$ ways to pick the first urn, and then $2$ ways to pick the second urn), and because $2$ balls first into urn $A$ and then $2$ balls into urn $B$ is the same if we swap urns $A$ and $B$, we need to divide by $2$.
So the answer is $\frac{3 \times 2}{2} = 3$ **ways**.
**Case 4**: $2$ balls in one urn, and one ball in each of the other urns. There are $3$ choices for the urn with $2$ balls. There is no other flexibility to choose, since one ball must occupy each of the other $2$ urns. So $3$ **ways** for this case.

Therefore there are $3 + 6 + 3 + 3 = 15$ **ways** to arrange the balls into urns.

## STARS AND BARS

Going through the casework for the simple case of 4 balls and 3 urns was time-consuming, and would become impossible to do without a computer as the numbers grow larger. There is, however, a technique that can be used to solve such problems. We will represent each ball by a star ★, and the boundary between urns by a vertical bar |. Any star to the left of the first bar is in urn $A$, any star between the pair of bars is in urn $B$, and any stars to the right of second bar is in urn $C$. So $\boxed{★ \mid ★★ \mid ★}$ can be read as 1 ball in urn $A$, 2 balls in urn $B$ and 1 ball in urn $C$. The 15 arrangements from the 4 cases on the previous slide are:

Case 1: $\boxed{★★★★ \mid\mid}$, $\boxed{\mid ★★★★ \mid}$, $\boxed{\mid\mid ★★★★}$.

Case 2: $\boxed{★★★ \mid ★ \mid}$, $\boxed{★★★ \mid\mid ★}$, $\boxed{★ \mid ★★★ \mid}$,

$\boxed{\mid ★★★ \mid ★}$, $\boxed{★ \mid\mid ★★★}$, $\boxed{\mid ★ \mid ★★★}$.

Case 3: $\boxed{★★ \mid ★★ \mid}$, $\boxed{★★ \mid\mid ★★}$, $\boxed{\mid ★★ \mid ★★}$.

Case 4: $\boxed{★★ \mid ★ \mid ★}$, $\boxed{★ \mid ★★ \mid ★}$, $\boxed{★ \mid ★ \mid ★★}$.

If we have $k$ balls and $n$ urns, we can represent them with a row of $k$ stars and $n-1$ bars ($n-1$ bars separate the stars into regions to represent the $n$ urns).

Using our combinatorics knowledge, we can figure out how many distinct ways there are to arrange a row of identical star and bar symbols. There are

$$\frac{(n+k-1)!}{k! \times (n-1)!}$$

distinct arrangements. This fraction is equivalent to $C_k^{n+k-1}$. We divide the $(n+k-1)!$ permutations by $k!$ because there are $k$ indistinguishable stars to arrange, and by $(n-1)!$ because there are $n-1$ identical bars to arrange. So for the 4 balls and 3 urns, there should be $C_4^6 = 15$ arrangements.

This was Problem 25 (the final and allegedly hardest) question on the AMC 8 in 2019. It is not really possible to solve this problem with casework due to time constraints on the exam, but seeing how we can apply "Stars and Bars" makes it computationally simple.

Alice has $24$ apples. In how many ways can she share them with Ben and Chris so that each of the three people has at least two apples?

The $24$ apples are identical in terms of counting, so they can be thought of as the "Stars". Each person is the destination for the apple, so they are the $3$ urns (meaning we have 2 "Bars"). There is an added complication, in that each person must have at least $2$ apples. No problem. Just subtract these $6$ apples from the $24$, and Alice must now split $18$ apples ($k = 18$) into $3$ groups ($n = 3$). This is $C_{18}^{20} = 190$.



Transforming a counting problem into a combinations calculation is possible here because we see that it's equivalent to arranging a pattern consisting of only two types of symbols: stars and bars.

1. What is the smallest possible average of four distinct positive even integers?

2. In a gymnastics competition, judge's marks are integers $0 - 10$ inclusive. There are $6$ judges, and the athlete's score is calculated by taking the average of the remaining $4$ marks after the highest and lowest marks are eliminated. The winning gymnast had a $8.5$ score, while the second place finisher had a $8.0$ score. If the both received identical lowest marks, could the highest marks they received have changed the order of finish if they were included in the scoring average? Explain your reasoning.

3. The *harmonic* mean of a set of non-zero numbers is the reciprocal of the average of the reciprocals of the numbers. What is the harmonic mean of $1$, $2$, and $4$?

4. What is median of the following data set: $2$, $3$, $3$, $4$, $4$, $4$?

5. In school's band string section, $28$ students play the violin, $18$ students play the viola, and $12$ students play both. How many students are in the string section?

6. There are $25$ students in a class. $12$ of them are on a football team and $14$ are on a soccer team. If $3$ students are on neither of these teams, how many students are on both the football and soccer teams?

7. How many integer pair $(x, y)$ solutions are there to $x^2 + y^2 < 10$?

8. If there are breakout rooms to work on $4$ different types of math problems, how many ways are there to split $20$ students into the breakout rooms, if each breakout room must contain at least $3$ students?

1. The Girl Scouts sold boxes of cookies outside of the school one morning. If they sold 252 boxes of cookies to 100 customers (and every customer bought at least one box of cookies), what is the maximum possible median number of boxes of cookies bought that morning?

2. In Graham Middle School's woodwind section, 21 students play clarinet, 16 students play oboe, 13 students play saxophone, 6 students play clarinet and oboe, 5 students play clarinet and saxophone, 4 students play oboe and saxophone. If there are 37 students total in the woodwind section, how many play all 3 instruments?

3. How many distinct positive integer triplets $(x, y, z)$ are solutions to $x + y + z = 12$?

4. When a fair 6-sided die is rolled 7 times, the probability that sum of the 7 rolls will be 10 can be represented as $\dfrac{n}{6^7}$. Find the value of $n$.