The Inclusion-Exclusion Principle
Introductory Example

Suppose a survey of 100 people asks if they have a cat or dog as a pet. The results are as follows:
55 answered yes for cat, 58 answered yes for dog and 20 people checked yes for both cat and dog. How many people have a cat or a dog?

Solution:

Since 55 have a cat, and 58 have a dog, you may think at first that $55 + 58 = 113$ have one or the other.

That thinking overlooks that some people – 20 of them – have both, and we counted these people twice by adding 55 and 58. To correct our answer, we must subtract from that sum the number 20:

$$55 + 58 - 20 = 93$$

have a cat or a dog. This is an example of the inclusion-exclusion principle.
The Case of Two Sets

Suppose we have two finite sets $A$ and $B$. Then what is the cardinality (number of elements) of $A \cup B$?

Again, if you take the sum of the individual cardinalities, then you count all elements in the intersection twice. Therefore, we have to subtract that cardinality to obtain the correct answer:

$$|A \cup B| = |A| + |B| - |A \cap B|$$

The generalization of this formula to an arbitrary number of sets is called the **inclusion-exclusion principle**.
The Case of Three Sets

Suppose we have three finite sets $A$, $B$ and $C$. Again, we ask for the cardinality of $A \cup B \cup C$.

We start with the provisional (incorrect) guess that it might be $|A| + |B| + |C|$. That sum counts some elements once, others more than once. Referring to the visualization on the right, it counts the elements in the pink areas once, the elements in the yellow areas twice, and the elements in the green center three times. To fix that, our next attempt is:

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

By subtracting the cardinalities of all the intersections of two sets, we ensure that “yellow” elements were counted only once.

However, the formula is still wrong for “green” elements. They were counted three times and then un-counted three times, for a net zero times. Therefore, the cardinality of the center needs to be added. The final (and correct) formula is:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$
The Inclusion-Exclusion Principle

The generalization of these formulas to an arbitrary number of sets is called the *inclusion-exclusion principle*. Given sets $A_1, A_2, \ldots, A_n$, the cardinality of the union is:

The sum of the individual cardinalities, minus all the cardinalities of intersections of two sets, plus the cardinalities of intersections of three sets, minus the cardinalities of intersections of four sets, etc. This alternating sum ends with plus or minus the cardinality of the intersection of all $n$ sets.

$$|A \cup B \cup C \cup D| = |A| + |B| + |C| + |D|$$
$$- |A \cap B| - |A \cap C| - |A \cap D| - |B \cap C| - |B \cap D| - |C \cap D|$$
$$+ |A \cap B \cap C| + |A \cap B \cap D| + |A \cap C \cap D| + |B \cap C \cap D|$$
$$- |A \cap B \cap C \cap D|.$$
The Case of Four Sets

When writing formulas like this, it is useful to check the number of terms of each kind using combinations:

There are $\binom{4}{2} = 6$ ways of picking two sets out of four sets without respect to order; therefore, there are 6 terms that are intersections of two sets.

There are $\binom{4}{3} = 4$ ways of picking three sets out of four sets without respect to order; therefore, there are 4 terms that are intersections of three sets.
A proof of the validity of the Inclusion-Exclusion Principle

Suppose the sets \( A_1, A_2, ... A_n \) are given. We wish to verify the correctness of the formula we gave two pages earlier.

Suppose \( x \) is an element in exactly \( k \) of the sets \( A_1, A_2, ... A_n \).

Then \( x \) was counted \( k \) times in the sum of the individual cardinalities.

It was then negatively counted \( C(k, 2) \) times in the course of subtracting all cardinalities of intersections of two sets.

It was then counted \( C(k, 3) \) times in the course of adding all cardinalities of intersections of two sets.

This continues until we get to \( C(k, k) \). Therefore, the total number of times \( x \) was counted is

\[
C(k, 1) - C(k, 2) + C(k, 3) - \cdots \pm C(k, k) = \sum_{j=1}^{k} C(k, j)(-1)^{j+1} = T
\]
The proof, continued

We learned previously that the alternating sum of all combinations for a given \( n \) is zero:

\[
0 = \sum_{k=0}^{n} C(n, k)(-1)^k
\]

Using this formula and replacing \( n \) by \( k \) and \( k \) by \( j \), we get

\[
0 = \sum_{j=0}^{k} C(k, j)(-1)^j
\]

Another way of writing this is

\[
0 = C(k, 0) + \sum_{j=1}^{k} C(k, j)(-1)^j
\]

By using \( C(k, 0) = 1 \) and moving the sigma sum to the left side, we get

\[
- \sum_{j=1}^{k} C(k, j)(-1)^j = \sum_{j=1}^{k} C(k, j)(-1)^{j+1} = T = 1
\]

Therefore, the number of times \( x \) is counted is exactly one, as it should be.
An inequality application of Inclusion-Exclusion

Problem: In a group of 50 people, 30 own a dog and 25 own a cat. What can we say about the number of people who own a dog and a cat?

Solution: let $A$ be the set of dog owners and $B$ be the set of cat owners. Then

$$|A \cup B| = |A| + |B| - |A \cap B|$$

We know $|A| = 30$, $|B| = 25$. We also know that $|A \cup B|$ can’t be greater than 50 because there are only 50 people total. Therefore,

$$50 \geq 30 + 25 - |A \cap B|$$

or

$$|A \cap B| \geq 30 + 25 - 50 = 5$$

Therefore, the answer is: at least 5 people must own both a dog and a cat.
An example with 3 sets

A survey in 1986 asked households whether they had a VCR, a CD player or cable TV.

40 had a VCR. 60 had a CD player; and 50 had cable TV.

25 owned VCR and CD player. 30 owned a CD player and had cable TV. 35 owned a VCR and had cable TV. 10 households had all three.

How many households had at least one of the three?

Solution: let $V$ be the set of households with a VCR. Let $C$ be the set of households with a CD player. Let $T$ be the set of households with cable TV.

The question is asking for $|V \cup C \cup T|$. By inclusion-exclusion, that is equal to

$$|V| + |C| + |T| - |V \cap C| - |V \cap T| - |C \cap T| + |V \cap C \cap T|$$

Therefore,

$$|V \cup C \cup T| = 40 + 60 + 50 - 25 - 30 - 35 + 10 = 70$$