Combinatorics Worksheet 3: Pigeonhole Principle

1. A teacher plans to give a five question true/false quiz to her class (300 students total). If any two students have the exact same answers, she will believe they are cheating. Explain why this plan is guaranteed to lead the teacher to accuse some students of cheating even if everyone is honest (of course there are many reasons that this plan is *likely* to lead the teacher to accuse honest students of cheating, but there is also a mathematical reason that the teacher is *guaranteed* to accuse some students of cheating).

> The number of ways a student can complete the quiz is 3^5: there are 5 questions and for each question they can either answer true or false or leave the question blank. Since $3^5 = 243$ is less than the number of students in the class, at least two students will complete the quiz in the same way.
> Formally, we are using the pigeonhole principle. The objects are the students and the boxes are the ways to complete the quiz. Since there are more objects than boxes, some box will end up with at least two objects.

2. Show that there are at least 250 four digit numbers whose digits all sum to the same value.

> The number of four digit numbers is $9 \cdot 10 \cdot 10 \cdot 10 = 9000$: there are 9 options for the first digit (since it cannot be 0) and 10 options for the remaining three digits.
> The number of possible sums of the digits of four digit numbers is 36: the lowest possible sum is 1, the highest possible sum is 36 (since that's the sum when every digit is 9), and all the possible sums are integers.
> Viewing the four digit numbers as objects and the possible sums as boxes, the pigeonhole principle implies that at least one box will end up with at least $\lceil \frac{9000}{36} \rceil = 250$ objects—in other words there is at least one value which is the digit sum of at least 250 four digit numbers.

3. Suppose that at a certain college there are only three majors: math, biology and CS. Also suppose that there are:

   - 85 students total
   - 50 math majors, 30 biology majors, and 40 CS majors
   - 10 double majors in math and biology, 20 in math and CS, and 10 in biology and CS.

   How many triple majors are there?

> The problem statement does not state whether triple majors also count as double majors. We will assume they do (in fact, if we didn't then the numbers given are impossible).
> Let M denote the set of math majors, B the set of biology majors, and C the set of computer science majors. By inclusion-exclusion, the total number of students is given

by

$$\begin{aligned} &|M| + |B| + |C| - |M \cap B| - |M \cap C| - |B \cap C| + |M \cap B \cap C| \\ &= 50 + 30 + 40 - 10 - 20 - 10 + |M \cap B \cap C| \\ &= 80 + |M \cap B \cap C|. \end{aligned}$$

Since we know that the total number of students is 85, we can find $|M \cap B \cap C|$ by solving

$$85 = 80 + |M \cap B \cap C|,$$

which gives $|M \cap B \cap C| = 5$. So the number of triple majors is $\boxed{5}$.

4. Show that there are more than 100 people alive right now who were born in the same minute.

For this problem, we need an estimate of the number of people alive right now. It is easy to find (reasonably credible) claims that the number is at least 7 billion.
Next we need to calculate the number of minutes in which people alive right now could have been born. The oldest known person alive today is younger than 120 years and there are $60 \cdot 24 \cdot 365.25$ minutes per year, so the maximum possible number of minutes in which people alive today could have been born is $120 \cdot 60 \cdot 24 \cdot 365.25 = 63113472$.
Viewing the people as objects and the minutes in which they could have been born as boxes and applying the pigeonhole principle, we see that there must have been some minute in which at least $\lceil \frac{7000000000}{63113472} \rceil = 111$ people were born.
In fact, we could improve this number a bit by leaving out very old people. There are not many people in the world over the age of 100, but leaving them out would allow us to significantly reduce the number of minutes we need to consider.

5. Show that at least 19 subsets of $\{1, 2, \ldots, 10\}$ have the same sum. For a challenge, try to improve this result as much as possible.

The number of subsets of $\{1, 2, \ldots, 10\}$ is 2^{10}.
The number of possible sums of subsets of $\{1, 2, \ldots, 10\}$ is 56: the lowest possible sum is 0 (here I am assuming that the sum of the empty set is 0, although it doesn't change the solution much if you just decide to leave out the empty set entirely), the largest possible sum is 55 (that's the sum of all ten numbers in the original set), and all the possible sums are integers.
Viewing the subsets as objects and the possible sums as boxes, the pigeonhole principle implies that at least one value is the sum of at least $\lceil \frac{2^{10}}{56} \rceil = 19$ different subsets.

6. **Challenge Problem:** Explain why there must be some files whose zipped versions are larger than the original files. Explain why the zip utility is called a compression algorithm despite this fact.