Problem Statement - Medicon Case Study

**Business Context**

Medicon is a leading pharmaceutical company and one of the world's largest vaccine manufacturers. During the last quarter, the company clinically tested five batches of the Covid-19 vaccine, and approximately 2,00,000 doses of this vaccine have already been administered to people. The sixth batch (40,000 doses) has now been manufactured and must be tested for quality assurance (whether the dose will be able to do a satisfactory job or not) as well as their time of effect (time taken for the dose to completely cure COVID). From the analysis of the previous batches, it is found that a dose is two times more likely to produce a satisfactory result as compared to an unsatisfactory result.

[Note: This is not a clinical trial, the vaccine has already cleared clinical trials. This batch test is to ensure the quality and effectiveness of the sixth batch]

**Objective**

By analyzing random samples collected from the batch, the company plans to infer the quality (whether a dose will do a satisfactory job or not) and time of effect of this (sixth) batch. As a Data Scientist in Medicon's Quality Assurance team, you've been tasked with inspecting the batch's quality. The objective has been sub-divided further into the following tasks:

**Task 1**: The Quality Assurance team has randomly selected 10 volunteers who were ready for the trial of the new batch and has given one dose to each volunteer. They want to understand the chances of doses being unsatisfactory by plotting the probability distribution of the number of unsatisfactory doses. They also want to find the probability that

A) exactly 3 doses out of 10 will not do a satisfactory job

B) at most 3 doses out of 10 will not do a satisfactory job

C) more than 8 doses out of 10 will do a satisfactory job

**Task 2**: The New York City administration wants to buy 20 doses for healthcare workers who tested positive for COVID. They have contacted the company and requested 20 doses. The Quality Assurance team want to find the probability that

A) at least 11 doses will not do a satisfactory job out of the 20 doses

B) at most 5 doses out of the 20 will not do a satisfactory job

C) at least 13 doses out of the 20 will do a satisfactory job

**Task 3**: The Quality Assurance team wants to analyze the time of effect for doses. So, they have randomly selected 50 volunteers with the help of the city administration and given one dose to each volunteer. The times of effect of these 50 doses are stored in the 'doses.csv' file. Based on the sample data, they want to

A) find the probability that the time of effect of a dose is less than 11.5 hours

B) find the probability that the time of effect of a dose is more than 10 hours

C) calculate the 90th percentile of time of effect for the doses, which will give the value below which lies the time of effect of 90% of the doses

**Note**: Here, satisfactory job means that the dose is able to completely cure covid and it does not have any side effects.

**Solution Approach**

Let LaTeX: \text{p} be the probability that a dose will not do a satisfactory job. It is provided that a dose is two times more likely to produce a satisfactory result as compared to an unsatisfactory result. So, the probability that a dose will do a satisfactory job is LaTeX: 2p.

According to the rules of probability, the total probability is always 1

LaTeX: p + 2p = 1

LaTeX:  \Rightarrow 3p = 1

LaTeX: \Rightarrow p = 1/3

LaTeX: \Rightarrow p = 0.33

Hence, the probability of success for each trial is 0.33.

**Note:** In this case, success denotes the event that a dose will not do a satisfactory job

**Data**

For the third task, you have been provided the dataset **doses.csv**. The features in the dataset are:

**drug\_serial\_number**: Unique serial number of a dose  
**time\_of\_effect**: Time taken for the dose to completely cure COVID (in hours)