

A 题 新冠疫苗接种问题研究

传染病具体指的是由各种病原体引起的能在人与人之间相互传播的一类疾病，所以传染病的传播需要具有三大要素：传染源、传播途径与易感人群，控制住任意一项都可以阻断传染病的传播。疫苗的作用就是保护易感人群。

天花是人们用疫苗消除的第一种传染病，它是由天花病毒感染人引起的一种烈性传染病，传染性强，死亡率极高。在人类消灭天花的过程当中，人们经过多年的探索和研究最终发明了牛痘疫苗。1976年，全球开始推行牛痘疫苗接种，实现了对所有易感人群的保护，天花病毒的感染被大大遏制，死亡率也有了明显的降低，进而成功阻断了天花的传播。1980年，世界卫生组织宣布根除天花。

不同传染病的传播力不一样，所以在人群当中要阻断其流行所需要形成的人群免疫水平也是不一样的。比如麻疹和百日咳的传染性较强，人群的免疫比例要达到90%以上才能阻断它们的流行，天花人群当中的免疫比例要达到80%以上才能阻断流行。

新冠肺炎的传染性较强，按照钟南山院士的说法，新冠肺炎的传播系数平均会达到3，有些地方甚至达到3.5，也就是说每个得病的人平均会有效传播3.5个人；而国外有些研究机构计算新冠肺炎的传播系数为2-2.5。国内外很多研究机构经过研究得出的结论，如果要想靠疫苗来阻断新冠肺炎的流行，人群免疫水平的阈值需要达到70%以上。

从官方发布的消息来看，我们正在接种的疫苗有效率在86%左右，而疫苗的有效期尚未清楚，世界卫生组织规定疫苗有效期必须超过半年，现在的疫苗基本上都没问题，但是最长能达到多少，由于疫苗上市时间不长，现在尚未有明确数据，美国瑞辉公司的疫苗，从已披露的数据来看，接种疫苗后三个月，从体内抗体水平可以推测应该有至少半年到一年的保护期，但这只是一个推测数据。美国传染病专家安东尼·福奇曾也表示疫苗可能无法提供长期免疫，预计有效期在3-6个月，最长不超过1年。

目前，在我们国家正在大力推进接种新冠疫苗的形势下，假设新冠肺炎的平均传播系数为 $r_0 = 3$ ，新冠肺炎的平均潜伏期平均为8天，疫苗的有效性约为85%，疫苗的有效期为6个月~12个月，而大部分人的有效期在9个月左右，其它所需

数据可自行网上查找，建立合适的传染病模型解决以下问题：

1. 需要多少比例的人接种疫苗才能阻断新冠肺炎的传播？
2. 结合目前我国国家的接种水平，预测我们国家构筑起免疫屏障需要的时间。
3. 在免疫屏障构建之后，长期来讲（3~5 年），需要如何动态接种疫苗，才能以最低的成本达到长期阻断新冠肺炎的传播的效果？

B 题 客户流失倾向问题研究

目前，电信运营商面临激烈的市场竞争，客户争夺愈演愈烈，用户流失是电信工业中最大的问题之一。传统意义上来讲，留住一个客户所需要的成本是争取一个新用户成本的 1/5，尤其对于剩余客户市场日渐稀疏的通信市场来说，减少客户流失就意味着用更少的成本减少利润的流失。然而问题是当运营商面临海量的客户资料时，应如何才能够从中提取出有效的信息以判断客户流失的状况或者倾向。因此对电信服务商的用户基础数据进行分析，探索对用户流失有深刻影响的相关指标，有助于帮助服务商制定有针对性的用户挽留计划，提高用户留存率。

本题为分类任务，即预测用户的流失情况（Churn 列为类别）。在给出的训练数据（train.csv）中，每个用户（一行代表一个用户）的前 19 维表示用户的一些特征（含义见数据说明），第 20 维（即 churn）代表该用户是否是流失客户，如果是流失客户对应 True，不是流失客户对应 False。

请完成如下两个问题：

(1) 使用 train.csv 数据构建用户特征与其是否是流失客户之间的模型，对测试数据 test.csv 进行预测（该数据包含了一些客户的 19 维特征，需要预测他们是否是流失客户），提交的答案：要求将预测结果替换文件 B01.csv 中的 class 列，并以自己的个人编号重命名该文件后提交。

(2) 衡量各因素对用户流失的影响。

附数据说明：

State	州代码
Account Length	帐户已启用多长时间
Area Code	区域代码
International plan	是否开通国际套餐
VMail Plan	是否开通语音信箱
Number vmail messages	收到的语音消息数量
Total day minutes	日间使用的总分钟数
Total day calls	日间通话总数
Total day charge	日间总消费
Total eve minutes	晚间使用的总分钟数
Total eve calls	晚间通话总数
Total eve charge	晚间总消费
Total night minutes	夜间使用的总分钟数
Total night calls	夜间通话总数
Total night charge	夜间总消费
Total intl minutes	总国际通话时间
Total intl calls	总国际通话数量
Total intl charge	总国际通话消费
Customer service calls	客户服务电话数量
Churn	是否流失