

# 2025 年第十五届 APMCM 亚太地区大学生数学建模竞赛

## B题 疾病的预测与大数据分析

为进一步加强以目标为导向的医疗质量安全管理工作，国家卫生健康委组织制定了《2025 年国家医疗质量安全改进目标》和 2025 年质控工作改进目标。

根据世界卫生组织（WHO）的数据统计，心血管疾病（CVD）是全球第一大死亡原因，估计每年夺去 1790 万人的生命，占全球死亡人数的 31%。附件中心脏病 heart.csv 数据集包含 11 个可用于预测可能的心脏病的特征。患有心血管疾病或心血管风险高的人（由于存在一种或多种危险因素，如高血压，糖尿病，高脂血症或已经确定的疾病）需要早期发现和管理。

此外，中风是全球第二大死亡原因，约占总死亡人数的 11%。本赛题附件中 stroke.csv 中风数据集中的每一行都提供了有关患者的相关信息，包含输入参数（如性别，年龄，各种疾病和吸烟状况）等指标，用于预测患者是否可能中风。第三种疾病肝硬化（cirrhosis）是由多种形式的肝病和病症（如肝炎和慢性酒精中毒）引起的肝脏瘢痕形成（纤维化）的晚期。

本次比赛提供了三种疾病数据集 stroke.csv、heart.csv 和 cirrhosis.csv，请你们团队运用数据统计与分析技能，深入挖掘数据信息，预测不同疾病发生的概率。

### 问题 1 数据预处理与基础统计分析

对三种疾病数据集 stroke.csv、heart.csv 和 cirrhosis.csv 进行数据预处理、统计分析和可视化，并分析哪些因素会影响中风、心脏病和肝硬化的患病概率。

### 问题 2 不同疾病预测模型的构建

请分别选取合适的特征指标，建立中风、心脏病和肝硬化三种疾病患病概率的预测模型，并进行模型准确性的检验、灵敏度分析和模型改进。

### 问题 3 多疾病关联与综合风险评估

请综合分析中风、心脏病和肝硬化这三种疾病的共同特征和共病情况，建立

数学模型预测同时患有其中任意两种和同时患有三种疾病的概率。

#### **问题 4** 预防三种疾病的建议和措施

请根据你们数学模型和数据分析的结果，针对这三种疾病，给世界卫生组织（WHO）写一封信，提出你们的预防建议和措施。