

2026MCM Problem C

本题是一个在具体背景下的数据分析及预测问题，与美赛一贯的重考察信息收集能力，尤其是在网络壁垒的条件下，美赛一般对该能力的考察更强。但是该题目将近乎所有所需信息与数据都给了我们，大大减少了我们信息收集的时间成本，适合数学能力及编程能力较强的同学。

问题一：开发一个或多个数学模型，估计每位选手在各周获得的观众投票数（真实数据未知且高度保密）。

- 你的模型是否能正确估计出导致真实淘汰结果的观众投票？请给出一致性指标。
- 你对估计结果有多大把握？这种不确定性是否在不同选手/周之间一致？请给出不确定性度量。

解答：

首先对数据进行预处理，本题中数据基本不存在异常值等情况，但是有缺失值（即 N/A），其中着重需要注意的是节目的评分标准的变化，依据题意，节目在前两季采用排名法，第三季到第 27 季采用百分比法，第 28 季淘汰机制进行了变动，并重新采用排名法一直到第 34 季。

针对第一第二季度我们对同一季度的几位选手提取出来，然后将它们的每周的评委得分相加，以此得到它们的评委分数总和，并根据此分数及最后的排名情况，估算观众排名，进而得到相应的观众票数。同样针对后续利用百分比法加和的季度，得到相应的评委票数百分比，与最后的排名情况结合，估算观众排名。

相应的模型可能性有很多，例如基准模型，线性模型，指数模型，帕累托模型，同情分模型，结合实际情况（粉丝最多的人往往其粉丝量远远多于粉丝第二多的人），估测指数模型更符合实际情况。通过代码计算其准确率，进而比较，得到最优模型及参数。

解答：

根据第一问的结果，第一小问的问题，即比较两种方法结果差异，能够非常简单的根据相应规则得到排名，通过比对即可得到结果。哪种方法更偏向观众意见，这一小问有多种思考方式，可以从理论上与实际结果上两方面得到。考虑实际结果方面可能更为准确与直观，即只考虑观众票数列出一个排名，将两种加和方法的排名分别与只考虑观众票数的排名进行对比，考虑差距最小的一个，即更偏向观众意见。

然后根据相应的指标分析争议选手，看其根据我们的评判方式与预估观众票数的相关信息，综合考虑。如果加入“评委在倒数两名中选择淘汰者”的机制，我们假设评委会选择评委分数较低的一方淘汰，根据此假设重新进入带入代码。考虑四人的最后成绩。

问题三：基于分析，你会推荐未来赛季使用哪种方法？是否建议保留评委二选一的机制？

利用数据（包括你估计的观众投票），建立模型分析专业舞者和明星特征（如年龄、行业等）对比赛结果的影响。这些因素对评委评分与观众投票的影响是否一致？

解答：

根据前面的考虑，为了减少争议与提升专业度，我们可以考虑评委比分权重较大的一种方式（与之对应，即上文得到的观众权重较小的一种方式），可以保留评委二选一机制加强专业性。将专业舞者与明星的年龄、行业按大类进行划分，例如电视圈，歌手，电影圈等。将其与我们得到的最后排名，评委评分与观众投票建模，分析他们的影响。可以考虑

问题四：设计一种你认为更“公平”或“更有观赏性”的评委与观众投票结合方式，并说明理由
解答：

综合前文，结合收视率与公平性考虑，我们可以综合采取排名法及评委分数阈值法。及先通过排名法加和并确定排名，其中评委分数过低（此处可设置阈值）的强制进入淘汰区（不论其排名，如果超过两名则优先分数最低的进入，以此类推），在此之外，最后选出两名综合排名最低的选手进入淘汰候选区，进一步这两位选手再进行一轮斗舞，由评委选择救回一组。这样既保证了公正，又保证了收视率。

参考文献：[1] Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323-351.

[2] Saari, D. G. (2000). Mathematical Structure of Voting Paradoxes: II. Positional Voting. *Economic Theory*, 15(1), 55-102.