

2026美赛 C 题 问题思路分析

这道题其实非常适合用**数据挖掘 + 逆向推断**的方法来做。如果你不熟悉博弈论、社会选择理论这些经济学模型，硬套反而容易露馅。用数据挖掘方法的话，核心逻辑就变成了：**基于历史淘汰结果的模式识别与投票逆推**。

背景知识：DWTS 投票机制

在动手之前，必须先搞清楚节目的投票规则。根据 Entertainment Weekly 的报道[1]和节目官方资料[2]：

评委评分：每周每对组合（明星+舞伴）表演后，由3-4位评委打分，每人满分10分。

观众投票：观众可以通过电话、短信、网站等方式投票。投票数据完全保密，从不公开！

淘汰规则：第1-2季用的是排名法（Ordinal Ranking），第3季之后改用百分比法（Percent Method）[2]。

方法	公式	说明
排名法	$\text{总排名} = \text{评委排名} + \text{投票排名}$	排名数字越小越好，相加后最大的被淘汰
百分比法	$\text{总比例} = 50\% \times \text{评委占比} + 50\% \times \text{投票占比}$	占比最低的被淘汰

关键争议：虽然官方说是50/50，但很多观众认为“观众投票权重更大”[3]。这就是问题二要分析的核心。

问题一：估计观众投票

问题抽象

问题一本质上是一个逆向推断问题（Inverse Problem）。

已知: 评委评分、每周淘汰结果
未知: 观众投票数/比例

用人话说就是: 我知道某人被淘汰了, 我知道他评委分是多少, 我反推他观众投票大概是多少。

重要提醒: 题目没有指定使用哪种投票合成方法!

历史上节目用过两种方法: 排名法 (S1-S2) 和百分比法 (S3起)。

所以问题一需要分别用两种方法建模, 得到两组投票估计!

为什么说这是“逆向问题”?

正向问题是: 已知投票 → 计算谁被淘汰。这很简单, 套公式就行。

逆向问题是: 已知谁被淘汰 → 反推投票。这就难了, 因为:

1. 解不唯一: 可能有无穷多组投票都能导致同一个淘汰结果
2. 信息损失: 淘汰结果只告诉你“谁最差”, 不告诉你其他人之间的差距
3. 噪声敏感: 评委分的小误差可能导致投票估计的大偏差

这类问题在科学领域很常见, 比如:

- 地震波数据 → 反推地下结构 (地球物理)
- CT扫描图像 → 反推人体组织 (医学成像)
- 观测轨迹 → 反推初始条件 (天体力学)

为什么选择这个视角?

很多人可能会想用回归/机器学习来预测投票, 但那需要有标签的训练数据——我恰恰没有! 投票数是保密的。

逆向推断的好处是: 它不需要投票的真实值, 只需要淘汰结果作为约束。这正好匹配我的数据情况。

建模思路

如果你熟悉运筹学, 这就是一个约束满足问题 (Constraint Satisfaction Problem)

核心策略: 针对排名法和百分比法, 分别构建两个优化模型

为什么要分别建模?

1. 两种方法的约束条件不同:
 - 排名法: 被淘汰者的 (评委排名 + 投票排名) 必须最大

获取更多数学建模相关资料关注【公众号：数模加油站】2026年美国大学生数学建模竞赛交流群：435813314

百分比法：被淘汰者的 $(50\% \text{评委占比} + 50\% \text{投票占比})$ 必须最低

2. 估计出的投票会不同：同样的淘汰结果，两种方法逆推出的投票分布可能差异很大

3. 这正是问题二需要的输入：问题二要对比两种方法，前提是问题一给出两组估

计两种方法的约束条件：

方法	综合得分公式	淘汰条件
排名法	$\text{Score}_j = \text{Rank_judge}(j) + \text{Rank_vote}(j)$	最大的被淘汰
百分比法	$\text{Score}_j = 0.5 \times \text{Pct_judge}(j) + 0.5 \times V_j$	最小的被淘汰

目标函数（在满足约束的前提下优化）：

因为解不唯一，我需要一个先验假设来挑选“最合理”的解。

最自然的假设是：观众倾向于给高分选手投票（虽然不完全一致，但大体正相关）。

所以目标函数设为： $\min \sum((V_j - V_{\hat{j}})^2)$ ，其中 $V_{\hat{j}}$ 是根据评委分预测的投票比例。这样做的好处：

1. 保证估计的投票一定能复现实际淘汰结果
2. 在多个可能解中选择“最不意外”的那个
3. 可以用成熟的凸优化求解器（cvxpy、ECOS等）高效求解
4. 两种方法分别建模，为问题二提供对比基础

用 Python + cvxpy 实现约束优化（两种方法）：

获取更多数学建模相关资料关注【公众号：数模加油站】2026年美国大学生数学建模竞赛交流群：435813314

对比维度	排名法	百分比法
信息使用	只用相对排名	使用绝对分数
分数差距	完全忽略	完全保留
优势选手	优势被压缩	优势被放大
冷门选手	更容易逆袭	很难逆袭

举个例子：

假设某周3位选手的评委分是：A=9分，B=8分，C=5分

用排名法：

- A排第1，B排第2，C排第3
- 评委分差距（9 vs 5 = 4分）被压缩成排名差距（1 vs 3 = 2个名次）

用百分比法：

- A占比 $9/(9+8+5) = 41\%$
- B占比 $8/(9+8+5) = 36\%$
- C占比 $5/(9+8+5) = 23\%$
- 差距保留：C要翻盘需要比A多18%的投票！

这意味着什么？

如果观众狂热支持C（比如C是偶像歌手有粉丝），用排名法C可能逆袭成功，用百分比法C很难翻盘。

所以节目组选择哪种方法，直接影响了“人气选手”能走多远。这就是问题二要分析的核心。

建模思路

数据输入：问题一估计的两组投票（votes_rank 和 votes_percent）

分析目标：这两种方法会导致多大程度的结果差异？在什么情况下差异最大？

步骤一：对每一季每一周，分别用排名法和百分比法计算综合排名

这是基础对比——看两种方法给出的淘汰人选是否一致。

步骤二：比较两种方法的淘汰结果是否不同

统计“不一致率”：多少周两种方法会淘汰不同的人？

问题三：影响因素分析

问题抽象

问题三其实就是一个回归分析问题——找出影响评委评分和观众投票的关键因素。

输入特征X：年龄、行业、来源地、舞伴

输出标签Y：评委分 / 观众投票 / 最终

名次这个问题的真正价值是什么？

表面上是找影响因素，实际上是要回答一个更深层的问题：

评委和观众的判断标准一样吗？

如果年龄对评委分影响大，但对观众投票影响小——说明评委看重“成熟稳重”，观众更喜欢“年轻活力”。

如果舞伴经验对两者影响都很大——说明舞伴选择对比赛结果影响巨大，可能需要在问题四里考虑。

为什么要从“因素分析”入手？

1. 可解释性强：机器学习预测准不准不重要，重要的是能解释“为什么”
2. 支撑问题四：设计新系统需要知道哪些因素最重要
3. 验证问题一：如果影响因素对评委分和估计投票的作用方向相反，说明估计是合理的建模思路

推荐方案：随机森林 + 特征重要性分析

为什么用随机森林而不是线性回归？

方法	优点	缺点
线性回归	系数直观	要求线性关系、特征独立
随机森林	能捕捉非线性、自动处理交互	黑箱，系数不好解释
XGBoost	精度高	更难解释

选择随机森林的理由：

1. 能处理分类特征（行业、舞伴）：不需要手动one-hot编码
2. 自动给出特征重要性排序：论文里直接用，说服力强
3. 不需要什么假设，简单粗暴：对数据分布没有要求
4. 能发现交互效应：比如“年轻运动员”和“年轻演员”可能表现不同

问题抽象

问题四其实是一个机制设计问题——设计一套规则，平衡不同利益相关者的诉求。

目标：设计一个既公平（专业性得到体现）又有趣（观众有参与感）的投票系统。

为什么这是一个难题？

公平性和娱乐性往往是矛盾的：

目标	极端情况	问题
完全公平	只看评委分	观众没有存在感，收视率下降
完全娱乐	只看投票	技术最差的可能夺冠，节目专业性受质疑

历史上DWTS就出现过“争议冠军”——评委分很低但粉丝投票把他抬到冠军。节目口碑受损。

所以我需要设计一个权衡机制，既尊重专业判断，又保持观众热情。

设计原则是什么？

1. 专业兜底：技术太差的选手不应该夺冠，再受欢迎也不行
2. 观众影响力：观众投票要有实质作用，不能沦为摆设
3. 悬念保持：比赛不能一眼看穿结果，要有逆转可能
4. 规则透明：观众能理解规则，知道自己投票有用

建模思路

三种方案，从简单到复杂：

方案一：动态权重机制

核心思想：随比赛进程调整权重

初期（第1-3周）：观众权重大（70%），增加参与感，让更多人关注

后期（第8-10周）：评委权重大（70%），保证冠军实至名归

为什么这样设计？

初期淘汰的都是“弱者”，让观众决定谁走其实影响不大

后期都是强者，需要专业眼光区分高下
渐进变化，观众不会感觉“突然被剥夺权力”

优点：简单、透明、容易解释

缺点：观众可能质疑“后期我的投票没用了”

根据EDA分析，以下发现对建模很重要：

发现	建模启示
评委分随周数递增（6.5→9.4分）	跨周比较要用相对排名，不能直接比分数
评委间高度一致（相关系数0.85+）	可直接用平均分代表评委意见
运动员和演员容易夺冠	行业是重要控制变量
Derek Hough参赛17季，6次冠军，胜率35%[4]	舞伴效应极其显著
部分赛季仅3位评委（S1-S8）	需分赛季处理评委数据

参考文献

- [1] Entertainment Weekly, "How does DWTS voting work? Executive producer explains"
<https://ew.com/>
- [2] Rice University, "Dancing with the Stars Voting Analysis"
<https://www.stat.rice.edu/~dobelman/dwts.html>
- [3] Collider, "DWTS voting controversy: Why audience votes seem to matter more"
<https://collider.com/>
- [4] Cheatsheet, "Derek Hough DWTS Statistics: 6 Wins in 17 Seasons"
<https://www.cheatsheet.com/>
- [5] Just Jared, "Former DWTS Pro Explains Voting System"
<https://www.justjared.com/>