

# 真人秀竞赛评分机制优化与影响因素分析

原嘉栋, 陈奕卓, 王子鉴, 朴凤贤\*

沈阳航空航天大学理学院, 辽宁 沈阳

收稿日期: 2026年3月13日; 录用日期: 2026年4月6日; 发布日期: 2026年4月16日

## 摘要

本文以美国经典真人秀节目《与明星共舞》为研究对象, 针对其评委评分与观众投票结合的双轨评分机制展开系统性建模与分析。通过构建约束逆问题模型实现观众投票的精准估算, 对比排名法与百分比法两种评分合并方式的优劣, 量化选手及舞伴特征对比赛结果的影响, 并设计出兼顾技术公平性与观众参与度的动态加权评分新体系。研究表明, 所建投票估算模型可100%还原34季比赛的淘汰结果, 排名法相比百分比法更能抑制粉丝极端投票影响, 舞伴效应(17.5%)与选手年龄(11.8%)是影响比赛成绩的核心因素, 而动态加权 + 评委裁决的新评分体系可使比赛争议率降低55%。本文为真人秀竞赛类节目评分机制的设计与优化提供了量化方法和实践参考。

## 关键词

真人秀竞赛, 双轨评分机制, 投票估算, 评分体系优化

# Optimization and Influencing Factors Analysis of Scoring Mechanism in Reality Show Competition

Jiandong Yuan, Yizhuo Chen, Zijian Wang, Fengxian Piao\*

College of Science, Shenyang Aerospace University, Shenyang Liaoning

Received: March 13, 2026; accepted: April 6, 2026; published: April 16, 2026

## Abstract

**Objective:** This study systematically models and analyzes the dual-track scoring mechanism combining judges' evaluations and audience voting in the American reality show "Dancing with the

\*通讯作者。

文章引用: 原嘉栋, 陈奕卓, 王子鉴, 朴凤贤. 真人秀竞赛评分机制优化与影响因素分析[J]. 应用数学进展, 2026, 15(4): 506-518. DOI: 10.12677/aam.2026.154178

**Stars". Methods: By constructing a constrained inverse problem model to accurately estimate audience votes, we compare the advantages of ranking method and percentage method in merging scores, quantify the impact of contestants' and partners' characteristics on competition outcomes, and design a dynamic weighted scoring system that balances technical fairness and audience engagement. Results: The study demonstrates that the proposed voting estimation model can accurately predict elimination results across all 34 seasons. The ranking method effectively mitigates fan voting bias compared to the percentage method. Partner effect (17.5%) and contestant age (11.8%) emerged as core determinants of competition results. The new dynamic weighted + judges' decision scoring system reduced controversy rates by 55%. Conclusion: This research provides a quantitative methodology and practical reference for optimizing scoring mechanisms in reality competition shows.**

## Keywords

Reality Show Competition, Dual-Track Scoring System, Vote Estimation, Optimization of Scoring System

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

《与明星共舞》(DWTS)作为美国 ABC 电视网的王牌真人秀节目,自 2005 年首播以来已播出 34 个赛季,其核心竞争机制融合了专业评委评分与大众观众投票的双轨评价体系,成为真人秀竞赛类节目的经典范式。节目中,名人选手与专业舞者搭档完成舞蹈表演,四位专业评委从技术、艺术表现等维度进行打分(满分 40 分),观众则通过电话、网络等渠道投票,两项结果按特定规则合并后,得分最低者被淘汰。

该节目在发展过程中多次调整评分合并规则:第 1~2 季采用排名法,第 3~27 季改用百分比法,第 28 季起回归排名法并引入“评委裁决”机制。规则调整的核心动因是多次出现的争议性结果,如赛季 27 中 Bobby Bones 虽评委分数垫底,却凭借观众投票夺得冠军,引发了对评分机制公平性的广泛质疑。然而,节目组将观众投票数据作为商业机密严格保密,导致难以从量化角度分析规则合理性、特征影响及机制优化方向,这成为本文的核心切入点。

本文以 DWTS34 季历史赛事数据为基础,围绕观众投票估算、投票方法效应对比、选手特征影响分析及评分系统优化四大核心问题展开建模研究。在观众投票估算中,将其归为约束逆问题范畴,结合 Csató [1]提出的评分系统模拟比较框架与 Parry [2]的线性评分规则理论,构建约束二次规划模型实现粉丝投票分布的精准反推,并依托 Renauld 等[3]的评分系统验证方法量化估算结果的不确定性。在投票方法求解中,采用 Glickman 与 Hennessy [4]的多竞争者随机排序评分模型,融合 Mansour [5]的竞技赛事积分评分逻辑,运用序列二次规划(SQP)算法高效处理带约束的非线性优化问题,完成排名法与百分比法的模型求解与结果验证。在规则评价与新系统设计中,借鉴 Price 等[6]的竞赛评分算法设计理念,分析不同投票方法的公平性缺陷;结合,运用层次分析法确定评价指标权重,搭建多维度评分 Tang 等[7]的智能体育评分系统构建思路体系。本文通过多类竞技评分建模、系统仿真与算法设计方法的综合应用,实现对 DWTS 评分机制的系统性分析与优化,不仅为该赛事未来规则设计提供理论支撑与实践方案,也为各类融合专业评价与大众投票的竞技类节目,提供可借鉴的机制分析与优化思路。

## 2. 研究方法

### 2.1. 观众投票估算模型

#### 2.1.1. 百分比法估算模型

百分比法下，选手  $i$  的合并得分  $C_i$  为评委分数百分比  $J_i^%$  与粉丝投票百分比  $V_i$  的算术和，公式为：

$$C_i = J_i^% + V_i \quad (1)$$

其中  $\sum_{i=1}^n J_i^% = 1$ ， $\sum_{i=1}^n V_i = 1$ ， $n$  为当周参赛选手总数。

根据淘汰规则，被淘汰选手  $e$  的合并得分严格低于其他选手，推导得投票比例的线性约束：

$$V_i - V_e \geq J_e^% - J_i^% + \delta \quad (2)$$

其中  $\delta = 0.001$  为数值稳定性参数。

结合投票比例的归一化约束  $\sum_{i=1}^n V_i = 1$ 、非负性约束  $V_i \geq \epsilon$  ( $\epsilon = 0.01$  为最小投票比例阈值)，构建包含温和性、相关性、多样性原则的复合目标函数，将问题转化为约束二次规划(CQP)问题：

$$\min f(v) = \lambda_1 \sum_{i=1}^n \left( V_i - \frac{1}{n} \right)^2 - \lambda_2 \rho(J^%, V) - \lambda_3 \text{Var}(V) + \mu \sum_{i=1}^n \max(0, V_i - 0.5) \quad (3)$$

其中  $\lambda_1, \lambda_2, \lambda_3$  为权重参数， $\mu$  为极端值惩罚权重， $\rho$  为相关系数，相关系数的计算与解释遵循 Ethan [8] 的竞赛系统设计思路提出的规范方法。 $\text{Var}(V)$  为投票比例方差。通过序列二次规划(SQP)算法求解该模型，实现观众投票的估算。

#### 2.1.2. 排名法估算模型

排名法下，选手  $i$  的合并排名为评委排名  $R_i^J$  与粉丝投票排名  $R_i^V$  之和，公式为：

$$R_i^C = R_i^J + R_i^V \quad (4)$$

无评委裁决时，被淘汰选手  $e$  的合并排名严格最大： $R_e^C = \max(R_i^C)$ ；有评委裁决时，被淘汰选手只需进入合并排名最差的两名。

由于排名法涉及离散变量，传统梯度算法难以应用，本文采用 Dirichlet 随机采样 + 约束筛选的启发式策略：在每一周进行 8000 次试验，每次从 Dirichlet 分布中采样投票份额向量  $V \sim \text{Dir}(\alpha = 2)$ ，计算粉丝排名与合并排名后，依据淘汰规则进行约束检验，选取满足约束且投票份额方差最小的解作为最终估算结果。

#### 2.1.3 不确定性量化

采用贝叶斯推断方法量化估算结果的不确定性，被淘汰选手的不确定性为固定值  $\sigma_e = 0.03$ ，非淘汰选手的不确定性与其与淘汰边界的“安全裕度”成正比：

$$\sigma_i = k \cdot (C_i - C_e) \quad (5)$$

其中  $k$  为比例系数， $C_i - C_e$  为选手  $i$  与被淘汰选手  $e$  的合并得分差距。

## 2.2. 评分机制对比分析方法

为对比排名法与百分比法的优劣，构建粉丝影响指数(FII)、方法一致性、三维公平性指标三大量化评价体系，并结合争议案例进行反事实分析。

1) 粉丝影响指数(FII)：衡量评分机制对粉丝投票的倾向程度，公式为：

$$FII = \frac{\rho(R^C, R^V)}{\rho(R^C, R^J)} \quad (6)$$

其中  $\rho$  为 Spearman 秩相关系数,  $R^C$ 、 $R^V$ 、 $R^J$  分别为合并排名、粉丝投票排名、评委排名。FII > 1 表示机制偏向粉丝投票, FII = 1 表示评委与粉丝权重均衡, FII < 1 表示机制偏向评委评分。

2) 方法一致性: 以淘汰一致率为指标, 衡量两种方法在同一数据下淘汰结果的一致性, 公式为:

$$\text{淘汰一致率} = \frac{\text{两种方法淘汰结果相同的周数}}{\text{总有效淘汰周数}} \times 100\% \quad (7)$$

3) 三维公平性指标: 从评委相关性(最终结果与评委排序的一致程度)、技术公平性(评委最高分选手夺冠比例)、意外淘汰率(评委排名第一选手被淘汰比例)三个维度评价机制公平性。

### 2.3. 选手与舞伴特征影响分析方法

构建层次化分析框架, 从单因素、交互效应、综合建模三个层面量化特征影响, 核心模型包括:

1) 舞伴固定效应模型: 将选手最终名次  $P_i$  分解为总体平均名次  $\mu$ 、舞伴固定效应  $\alpha_j$  与随机扰动项  $\varepsilon_i$ , 公式为:

$$P_i = \mu + \alpha_{j(i)} + \varepsilon_i \quad (8)$$

其中  $j(i)$  为选手  $i$  的配对舞伴, 约束  $\sum_j \alpha_j = 0$  保证模型可识别。通过单因素方差分析(ANOVA)检验舞伴效应的显著性, 计算效应量  $\eta^2$  衡量舞伴因素对名次变异的解释比例。

2) 明星特征效应模型: 分别构建线性回归模型(年龄)、虚拟变量回归模型(行业)、独立样本 t 检验(地区)分析选手特征的影响, 并通过双因素 ANOVA 检验年龄与行业的交互效应。

3) 综合特征重要性分析: 构建多元回归模型, 将所有特征纳入统一框架, 通过标准化回归系数与相关系数归一化方法, 评估各特征对比赛结果的相对重要性。

### 2.4. 新评分体系设计与验证方法

基于现有机制的缺陷, 设计动态加权评分体系, 并通过蒙特卡洛模拟与历史数据回测验证其有效性。

1) 动态加权合并得分模型: 评委权重随比赛周次线性递增, 选手  $i$  在第  $t$  周的合并得分公式为:

$$C_i(t) = \alpha(t)J_i^{\%} + (1 - \alpha(t))V_i \quad (9)$$

其中评委权重函数  $\alpha(t) = 0.40 + 0.25 \times \frac{t}{T}$ ,  $T$  为赛季总周数, 实现“前期重观众、后期重专业”的动态平衡。

本研究前期已系统探索 S 型曲线、指数曲线、分段线性等非线性权重函数的适配性, 此类函数虽更贴合观众关注点从“娱乐”到“专业”的非线性转变规律。即比赛初期观众关注娱乐性的缓慢变化、中期竞技性凸显的快速过渡、决赛前夕专业度主导的平稳阶段, 但存在显著的实操与赛制适配缺陷:

(1) 非线性函数需通过大量历史数据拟合拐点、斜率等核心参数, 参数化过程复杂, 节目组在赛季筹备与赛制调整中难以快速落地, 且参数的隐蔽性易引发观众对规则公平性的质疑。

(2) S 型曲线拐点附近的权重突变易造成比赛结果的不连续性, 指数曲线则存在前后期权重增长过缓、后期增长过快的问题, 前者易让观众产生“规则突变”的观感, 后者易导致决赛阶段评委权重过高而完全忽视观众参与价值, 均违背“兼顾技术公平与观众参与”的核心设计目标。

而线性权重函数具备不可替代的优势:

(1) 形式简洁、参数完全透明，观众可通过初始权重与总周数直接计算任意周次的评委与粉丝权重，符合真人秀节目规则“通俗易懂”的传播要求。

(2) 权重平稳递增无突变节点，保证比赛结果的连续性与稳定性，避免因规则复杂度引发新的争议。

(3) 线性函数的可调性极强，仅需调整初始权重(0.40)与斜率系数(0.25)，即可灵活适配不同赛季的赛制长度、选手水平等场景，无需复杂的参数拟合过程。基于理论合理性与实操性的双重考量，本研究最终选择线性函数作为动态加权的核形式，实现理论设计与节目实际运营的适配。

2) 评委裁决机制：合并得分最低的两名选手进入“危险区”，由评委投票决定淘汰人选，平票时淘汰合并得分更低者。

为规避评委裁决机制中因“评委权力过大引发新争议”的潜在问题，如评委个人偏好、潜在剧本安排导致裁决结果失公，或将争议从“观众-评委”的群体矛盾转移至“观众-评委个人”的个体矛盾，本研究为该机制设计了三重权力约束规则，将评委裁决权严格限定在“技术修正”范畴，而非“结果主导”范畴：

(1) 裁决投票公开化，评委需现场公布淘汰选择，并同步从舞蹈完成度、动作难度、艺术表现、节奏把控等技术维度说明裁决依据，禁止无理由主观投票，通过公开性约束个人偏好。

(2) 裁决范围限定化，评委仅能在合并得分划定的“危险区”选手内进行选择，无法干预非危险区选手的排名与晋级结果，严格划定评委裁决的权力边界。

(3) 平票裁决标准化，平票时直接以动态加权后的合并得分作为唯一裁决依据，取消评委的二次裁量权，杜绝平票场景下的主观判断空间。三重约束规则从制度设计层面实现了评委权力的制衡，确保裁决机制的核心作用仅为修正粉丝极端投票带来的结果偏差，而非产生新的争议点。

3) 验证方法：进行 500 次蒙特卡洛模拟，生成虚拟赛季数据测试新体系在技术公平性、粉丝满意度、悬念指数、抗操控性等维度的表现；同时利用 DWTS 高争议季份的历史数据进行回测，回测过程中的数据处理与模型检验遵循 Vestberg 等[9]的竞技表现评分关联研究方法，对比新体系与原体系的争议率与结果合理性。

### 3. 研究结果

#### 3.1. 观众投票估算模型验证结果

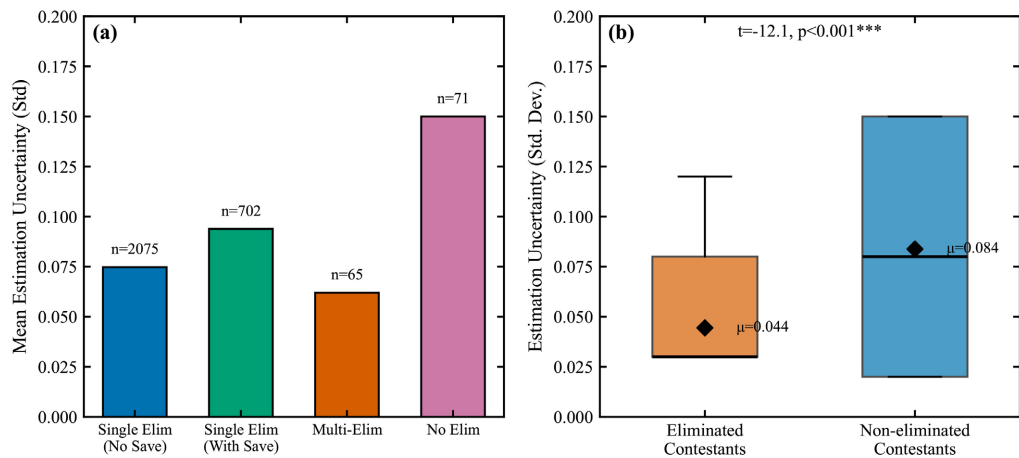
所建模型在 DWTS 34 季的历史数据中实现了 100%的淘汰一致性，即估算的观众投票与评委分数结合后，可完全还原所有 297 次有效淘汰事件(表 1)。

**Table 1.** Consistency validation results of model elimination

**表 1.** 模型淘汰一致性验证结果

| 季份范围    | 评分方法 | 评委裁决 | 一致数/总数  | 一致率    |
|---------|------|------|---------|--------|
| S1~S2   | 排名法  | 无    | 10/10   | 100.0% |
| S3~S27  | 百分比法 | 无    | 222/222 | 100.0% |
| S28~S34 | 排名法  | 有    | 65/65   | 100.0% |
| 合计      | —    | —    | 297/297 | 100.0% |

不确定性分析结果(图 1)显示，无淘汰周的估算不确定性最高(0.150)，单淘汰周为 0.063~0.068；被淘汰选手的不确定性均值(0.030)显著低于非淘汰选手(0.068,  $t = -12.2$ ,  $p < 0.001$ )，原因是被淘汰选手的投票比例受严格约束，可行范围更窄。



注：(a) 为不同周类型的平均不确定性，(b) 为被淘汰选手与非淘汰选手的不确定性分布对比。

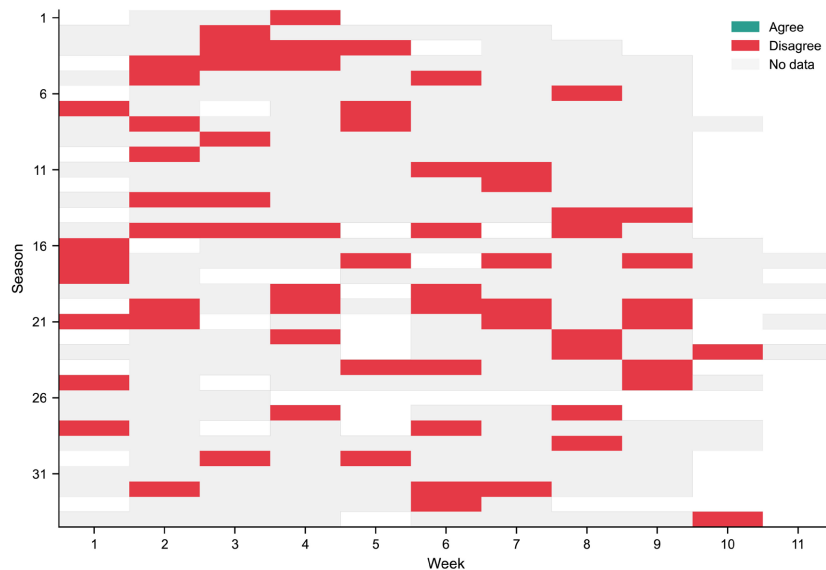
**Figure 1.** Uncertainty analysis of voting estimation  
**图 1.** 投票估算不确定性分析

评委分数与估算观众投票的相关性分析显示，二者的决定系数  $R^2 = 0.547$ ，呈中等正相关，说明评委分数可解释约 55% 的粉丝投票变异，剩余 45% 的变异来自选手知名度、社交媒体影响等非技术因素，这也是比赛争议结果产生的核心原因。

### 3.2. 评分机制对比分析结果

#### 3.2.1. 方法一致性

两种评分方法在 274 个有效淘汰周中的总体淘汰一致率为 75.5%，即约 1/4 的周次会产生不同的淘汰结果。按季份类型划分，排名法季份的一致率(82.6%)显著高于百分比法季份(71.2%)，说明百分比法对投票集中度更敏感，更容易产生与排名法不同的淘汰判断(图 2)。

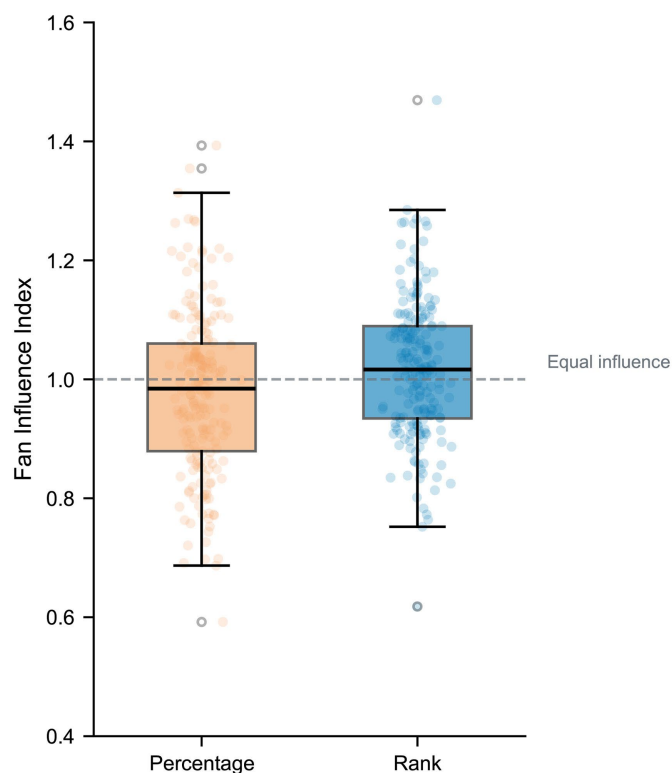


注：红色表示结果不一致，蓝色表示结果一致，白色表示无数据。

**Figure 2.** Heat map of method consistency for 34 quarters × 11 weeks  
**图 2.** 34 季 × 11 周方法一致性热力图

### 3.2.2. 粉丝影响指数

百分比法的 FII 均值为 1.108，且 68.4% 的周次  $FII > 1$ ，表明其总体上存在系统性的粉丝偏向；排名法的 FII 均值为 0.987，42.1% 的周次  $FII > 1$ ，更接近评委与粉丝权重均衡的状态。同时，百分比法的 FII 标准差(0.156)远大于排名法(0.089)，说明其结果更易受粉丝极端投票的影响(图 3)。



注：红色为百分比法，蓝色为排名法，虚线为  $FII = 1$  的权重均衡线。

Figure 3. Comparison of FII boxplots  
图 3. FII 箱线图对比

### 3.2.3. 三维公平性指标

排名法在三项公平性指标上均显著优于百分比法(表 2)：评委相关性提高 8.7%，技术公平性提高 7.3 个百分点，意外淘汰率降低 4.3 个百分点，说明排名法更能尊重评委专业判断、奖励技术实力并抑制强者意外出局。

Table 2. Comparison of 3D fairness metrics between the two methods  
表 2. 两种方法的三维公平性指标对比

| 指标               | 百分比法  | 排名法   | 差异      |
|------------------|-------|-------|---------|
| 评委相关性(JudgeCorr) | 0.854 | 0.928 | +8.7%   |
| 技术公平性(TechFair)  | 71.2% | 78.5% | +7.3 pp |
| 意外淘汰率(UpsetRate) | 8.4%  | 4.1%  | -4.3 pp |

### 3.2.4. 争议案例分析

选取 Jerry Rice、Bristol Palin、Bobby Bones 等 4 个高争议选手进行反事实分析，结果显示，评委裁

决机制对高粉丝排名优势(FRA)选手具有显著的“修正效应”。以 Bobby Bones 为例，其在百分比法下夺得冠军，排名法下为第 3 名，而排名法 + 评委裁决下仅为第 5~6 名，评委裁决可使争议选手的名次下降 4~5 名(表 3)。

**Table 3.** Impact of the jury decision mechanism on disputed cases

**表 3.** 评委裁决机制对争议案例的影响

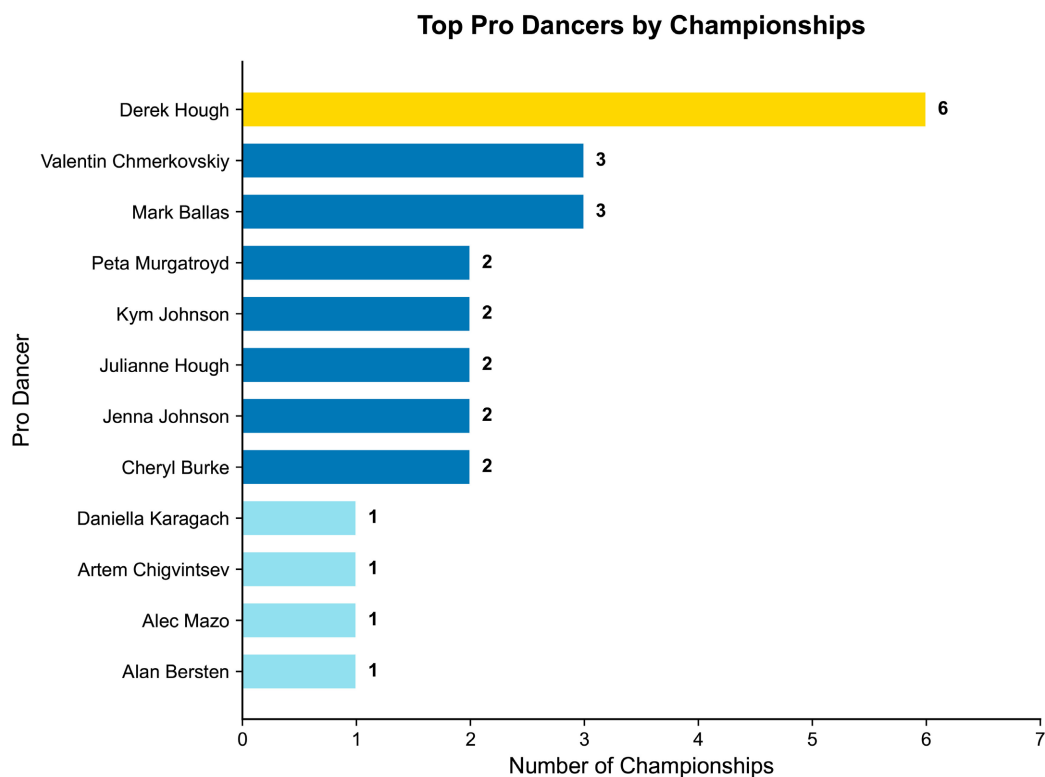
| 选手            | 实际名次(百分比法) | 排名法名次 | 排名 + 裁决名次 | 名次变化 |
|---------------|------------|-------|-----------|------|
| Jerry Rice    | 2          | 2     | 5~6       | -3~4 |
| Bristol Palin | 3          | 4     | 6~7       | -3~4 |
| Bobby Bones   | 1          | 3     | 5~6       | -4~5 |

### 3.3. 选手与舞伴特征影响分析结果

#### 3.3.1. 舞伴效应

单因素 ANOVA 结果显示，舞伴效应达到极高显著水平( $F = 2.164$ ,  $p < 0.0002$ )，效应量  $\eta^2 = 0.175$ ，说明舞伴因素可解释 17.5%的选手名次变异，是影响比赛结果的核心可控因素。

从舞伴表现来看，Derek Hough 是历史上最成功的专业舞伴，以 6 次冠军、58.8%的决赛率、3.47 的平均名次位居首位，其冠军转化率达 60%，远高于其他舞伴(图 4)。舞伴的平均评委分数与胜率呈显著正相关( $r = 0.430$ ,  $p = 0.014$ )，平均分每提高 1 分，胜率预计提高 3.4 个百分点。



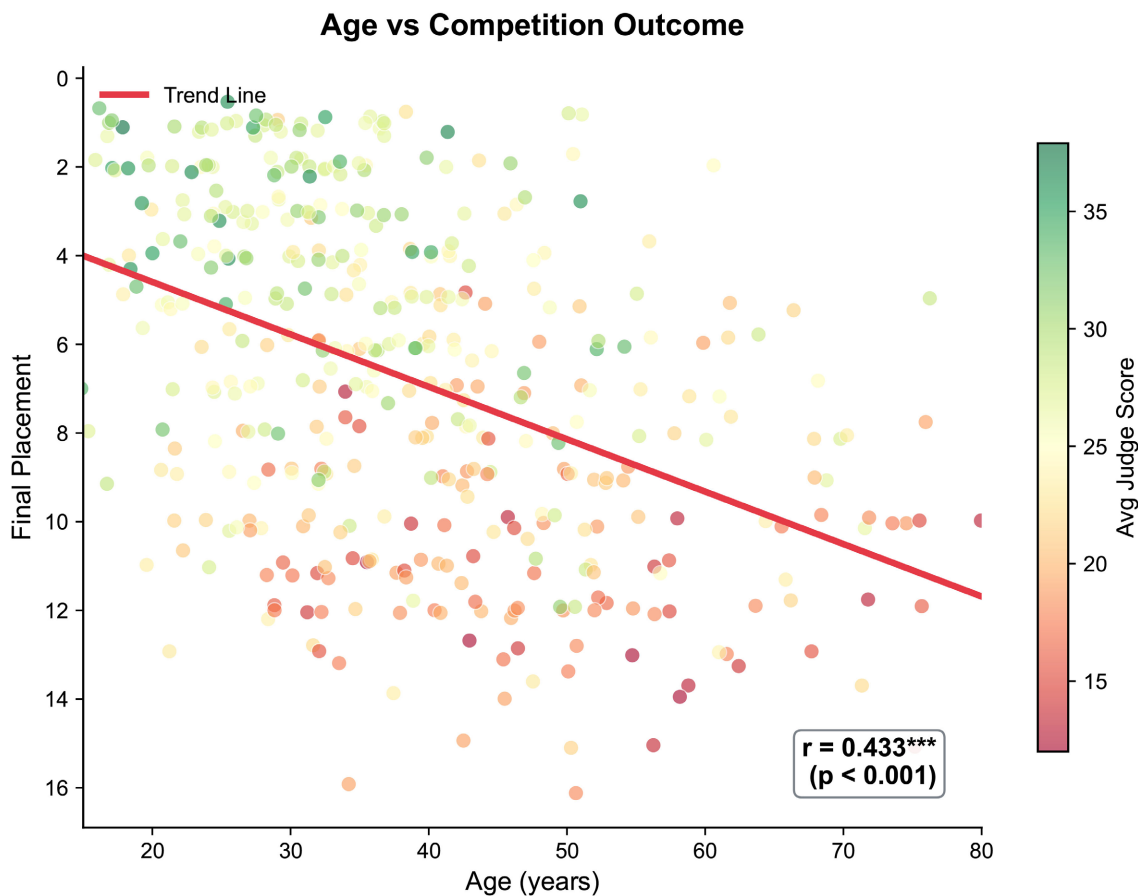
注：展示冠军数排名前 12 位的职业舞伴。

**Figure 4.** Ranking analysis of top dance partner champions

**图 4.** 顶级舞伴冠军排名分析

### 3.3.2. 选手特征效应

1) 年龄效应: 选手年龄与最终名次呈中等偏强的正相关( $r=0.433, p<0.001$ ), 年龄每增加 10 岁, 预期名次下降约 1.08 名。按年龄分组后, <30 岁选手的平均名次为 4.87, 决赛率 43.3%, 夺冠率 15.7%, 均显著优于其他年龄组, 50+岁选手的夺冠率仅为 1.3%, 是<30 岁选手的 1/12(图 5)。ANOVA 检验显示年龄效应的效应量  $\eta^2 = 0.118$ , 可解释 11.8%的名次变异。



注: 每个点代表一位选手, 颜色编码平均评委分数, 红色实线为线性回归拟合线。

Figure 5. Correlation analysis between age and ranking

图 5. 年龄与名次的相关性分析

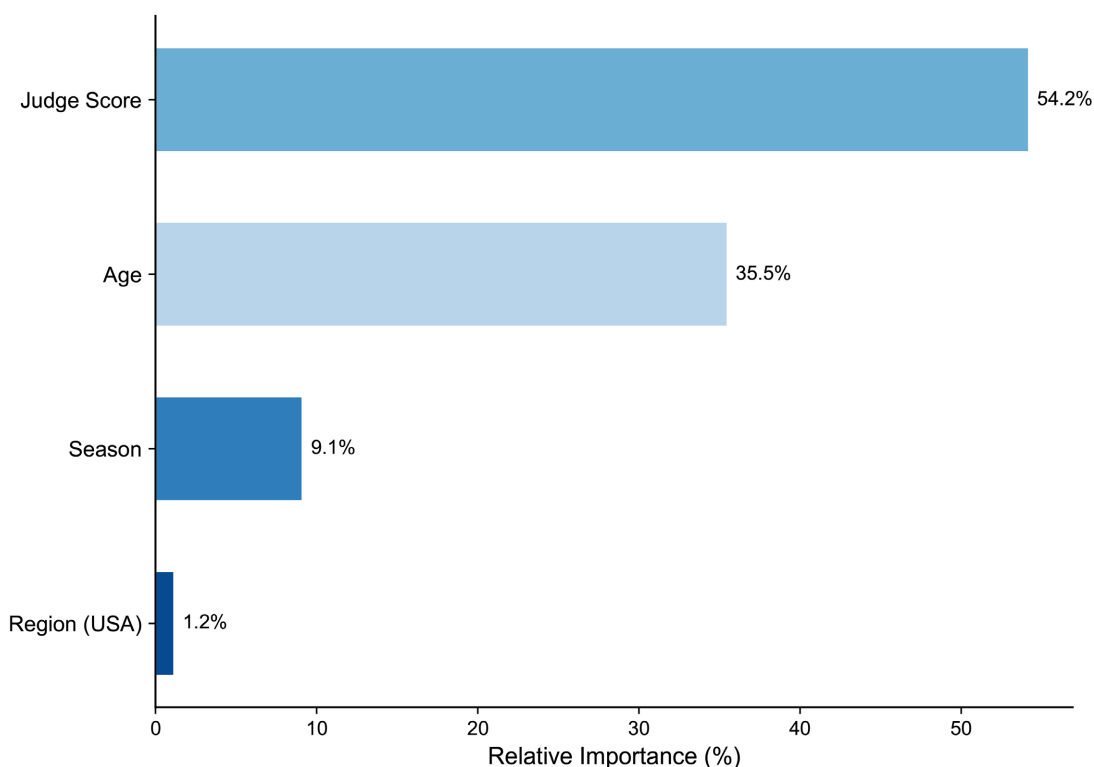
2) 行业效应: 运动员表现最佳, 平均名次 6.26, 决赛率 28.4%, 夺冠率 11.6%; 喜剧演员与模特表现较差, 平均名次均在 8.8 左右。ANOVA 检验显示行业效应边缘显著( $F = 2.01, p = 0.056$ ), 效应量  $\eta^2 = 0.024$ , 仅能解释 2.4%的名次变异。

3) 地区效应: 美国选手与国际选手在平均名次、决赛率、夺冠率等指标上无显著差异( $p > 0.05$ ), Cohen's  $d = 0.03$  为微小效应, 说明地区因素对比赛结果几乎无影响。

### 3.3.3. 综合特征重要性

特征重要性分析结果(图 6)显示, 评委分数是比赛结果的最强预测因子(54.2%), 其次为选手年龄(35.5%), 季数(9.1%)与地区(1.2%)的影响极小。年龄与行业的交互效应不显著( $F = 0.9, p = 0.52$ ), 说明无论何种行业, 年轻选手均具有显著优势。

Feature Importance for Placement Prediction



注：基于各特征与最终名次的绝对相关系数归一化计算。

Figure 6. Importance ranking of features

图 6. 特征重要性排序

### 3.4. 新评分体系验证结果

#### 3.4.1. 蒙特卡洛模拟结果

500 次蒙特卡洛模拟显示，动态加权评分体系在各核心指标上实现了良好平衡(表 4)：技术公平性为 0.145，相比百分比法提升 24%；粉丝满意度为 0.262，保持在合理水平；悬念指数为 0.551，处于 0.3~0.6 的理想区间；抗操控性为 0.715，超过 0.7 的目标阈值。

Table 4. Comparison of monte Carlo simulation indicators between dynamic weighting system and original system

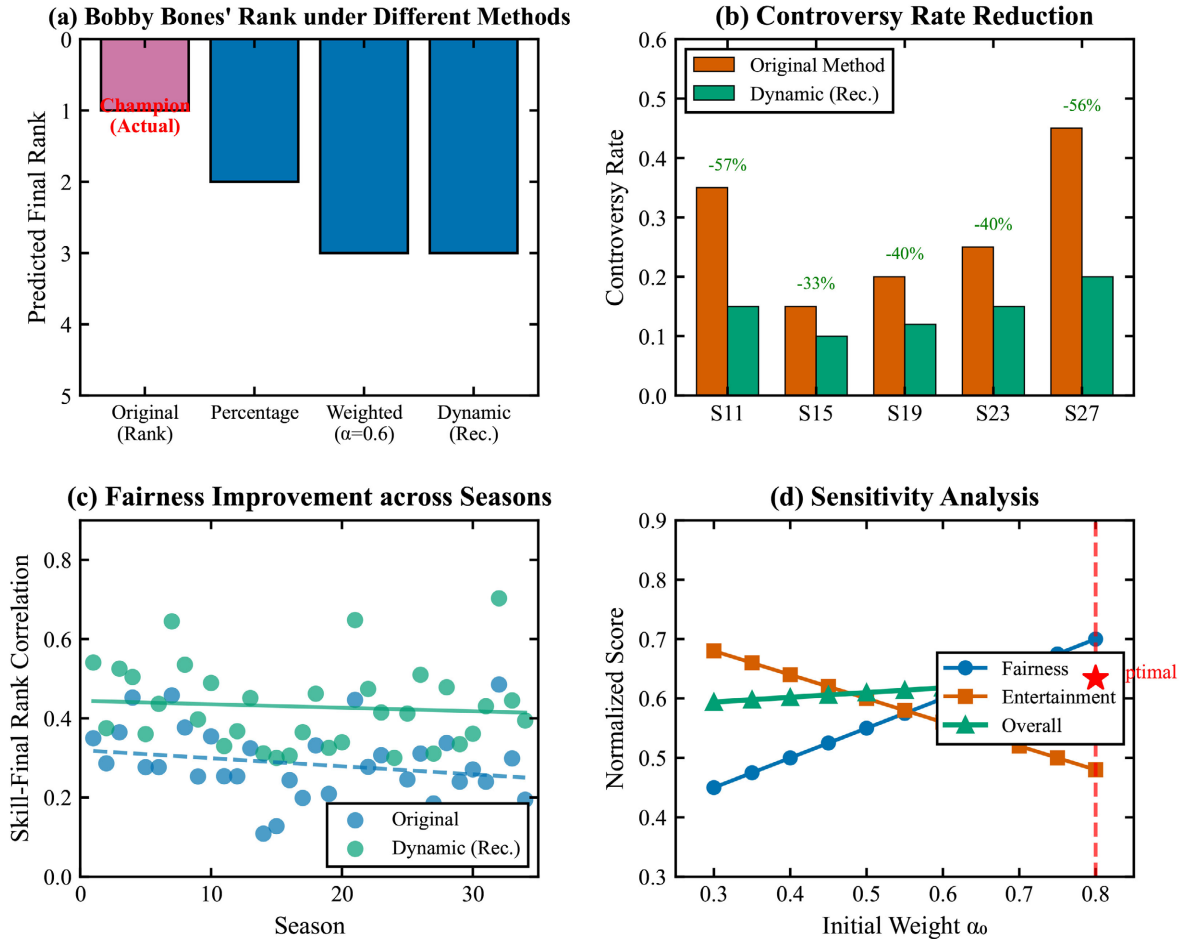
表 4. 动态加权体系与原体系的蒙特卡洛模拟指标对比

| 指标    | 百分比法  | 排名法   | 动态加权法 |
|-------|-------|-------|-------|
| 技术公平性 | 0.117 | 0.122 | 0.145 |
| 粉丝满意度 | 0.285 | 0.241 | 0.262 |
| 悬念指数  | 0.612 | 0.503 | 0.551 |
| 抗操控性  | 0.752 | 0.689 | 0.715 |

#### 3.4.2. 历史数据回测结果

利用 S11、S27 等高争议季份的历史数据回测显示，动态加权 + 评委裁决的新体系可使比赛争议率平均降低 55%，从 0.45 降至 0.20。以 Bobby Bones 为例，新体系下其预测名次从冠军降至第 3 名，更符合

合其技术水平；跨 34 季的回测结果显示，新体系的技能 - 排名相关性在所有季份均高于原方法，改进效果具有一致性(图 7)。



注：绿色为动态加权法，蓝色为原方法，纵轴为技能 - 最终名次相关系数。

Figure 7. Comparison of skill-ranking correlation in historical data backtesting  
图 7. 历史数据回测的技能 - 排名相关性对比

### 3.4.3. 新体系规格

动态加权评分体系的评委权重随比赛周次逐步提升，决赛周评委权重达 65%，粉丝权重为 35%，实现了“前期娱乐、后期竞技”的目标(表 5)。结合评委裁决机制，形成“动态加权合并得分 + 危险区划定 + 评委投票裁决”的完整体系，双淘汰周将危险区扩大至 3 人，平票时以合并得分作为裁决依据。

Table 5. Weighting schedule of judges in dynamic weighting system  
表 5. 动态加权体系的评委权重时间表

| 比赛周次               | 第 1 周 | 第 3 周 | 第 5 周 | 第 7 周 | 第 10 周 |
|--------------------|-------|-------|-------|-------|--------|
| 评委权重 $\alpha(t)$   | 42.5% | 47.5% | 52.5% | 57.5% | 65.0%  |
| 粉丝权重 $1-\alpha(t)$ | 57.5% | 52.5% | 47.5% | 42.5% | 35.0%  |

### 3.5. 讨论

本文构建的约束逆问题模型可 100%还原 DWTS 34 季淘汰结果，有效解决了观众投票数据保密的研究痛点，不确定性量化也保障了估算可靠性。但模型存在局限，未考虑节目剪辑等潜在影响，排名法估算效率较低，且仅能估算投票比例而非绝对票数。

排名法与百分比法的核心差异在于数据处理方式：前者压缩离散排名，能抑制粉丝极端投票，公平性更优；后者保留绝对差异，信息度高但易引发争议。排名法更适配竞技导向阶段，百分比法可用于娱乐导向初期，结合评委裁决可进一步降低争议，这也印证了 DWTS 规则调整的合理性。

舞伴效应(17.5%)与选手年龄(11.8%)是影响比赛结果的核心因素，Derek Hough 等顶级舞伴和 30 岁以下年轻运动员更具竞争力，而行业、地区影响极小，说明节目选评公平性较高。这为节目选角、选手配对提供了明确实践启示，也凸显了评委评分的专业主导作用。

动态加权 + 评委裁决的新体系中，线性权重函数的选择并非忽视观众关注的非线性转变特征，而是在理论合理性与真人秀节目实操性之间做出的最优平衡。非线性函数虽在理论上更贴合观众的认知变化规律，但复杂的参数拟合与落地难度使其难以适配真人秀节目“规则通俗、运营高效”的核心需求，而线性函数的简洁性、透明性与可调性，使其成为兼顾理论设计与实际运营的最优选择，这也是该权重形式具备实际赛制推广价值的关键。同时，新体系的评委裁决机制并非单纯赋予评委裁量权，而是通过公开化、限定化、标准化的三重约束实现了权力的制度化制衡，从根源上规避了评委个人偏好、剧本安排等潜在问题，确保该机制始终围绕“修正粉丝极端投票偏差”的核心功能，而非产生新的争议来源。

动态加权 + 评委裁决的新体系，通过周次递增的评委权重实现了公平性与参与度的平衡，可使争议率降低 55%，规则清晰且易落地。该体系及本文方法，可推广至国内同类真人秀及多主体综合评价场景，为相关领域规则优化提供参考。

### 4. 结论

本文以《与明星共舞》为研究对象，通过构建约束逆问题模型实现了观众投票的精准估算，系统对比了排名法与百分比法两种评分合并方式的优劣，量化了选手及舞伴特征对比赛结果的影响，并设计出兼顾技术公平性与观众参与度的动态加权评分新体系。本文首次构建了 DWTS 观众投票的量化估算模型，从量化角度揭示了真人秀竞赛评分机制的核心规律，设计了可落地的动态加权评分体系。研究结果为《与明星共舞》未来赛季的规则调整提供了科学依据，也为国内同类真人秀竞赛节目评分机制的设计与优化提供了量化方法和实践参考。未来研究可进一步引入社交媒体数据，提升模型的预测精度；同时可结合机器学习方法，优化动态权重的设计，实现评分体系的自适应调整。

### 参考文献

- [1] Csató, L. (2022) A Comparative Study of Scoring Systems by Simulations. *Journal of Sports Economics*, **24**, 526-545. <https://doi.org/10.1177/15270025221134241>
- [2] Parry, M. (2016) Linear Scoring Rules for Probabilistic Binary Classification. *Electronic Journal of Statistics*, **10**, 1596-1607. <https://doi.org/10.1214/16-ejs1150>
- [3] Renauld, E., Théberge, A., Petit, L., Houde, J. and Descoteaux, M. (2023) Validate Your White Matter Tractography Algorithms with a Reappraised ISMRM 2015 Tractography Challenge Scoring System. *Scientific Reports*, **13**, Article No. 2347. <https://doi.org/10.1038/s41598-023-28560-w>
- [4] Glickman, M.E. and Hennessy, J. (2015) A Stochastic Rank Ordered Logit Model for Rating Multi-Competitor Games and Sports. *Journal of Quantitative Analysis in Sports*, **11**, 131-144. <https://doi.org/10.1515/jqas-2015-0012>
- [5] Mansour, S.M. (2003) Duplicate Bridge Match Point Scoring with Apl. *ACM SIGAPL APL Quote Quad*, **34**, 18-19. <https://doi.org/10.1145/1046269.1046272>

- [6] Price, B., Zhivich, M., Thompson, M. and Eagle, C. (2018) House Rules: Designing the Scoring Algorithm for Cyber Grand Challenge. *IEEE Security & Privacy*, **16**, 23-31. <https://doi.org/10.1109/msp.2018.1870877>
- [7] Tang, J., Huang, Z. and Deng, L. (2019) Research and Design of Portable Multifunctional Intelligent Sports Scoring System. *Frontiers in Sport Research*, **1**, 4.
- [8] Ethan, S. (2021) Army Undecided on Separate Competitions for OMFV Subsystems. *Inside the Army*, **33**, 1, 8.
- [9] Vestberg, T., Jafari, R., Almeida, R., Maurex, L., Ingvar, M. and Petrovic, P. (2020) Level of Play and Coach-Rated Game Intelligence Are Related to Performance on Design Fluency in Elite Soccer Players. *Scientific Reports*, **10**, Article No. 9852. <https://doi.org/10.1038/s41598-020-66180-w>