1 数据处理与特征构建

1.1 数据结构与处理目标

原始数据由两部分组成：其中第一部分是静态数据，由“选手–赛季”给出（如 celebrity_name、ballroom_partner、celebrity_industry、celebrity_homestate、

celebrity_homecountry_region、celebrity_age_during_season、results、placement 等）；第二部分是按周记录的评委的打分，列名为 weekX_judgeY_score。我们需要先估计每位选手每周的粉丝投票数才能够进行问题的探讨，因此我们先把数据整理为“选手–周”面板表（每行对应一个选手在某赛季某周的记录），并在此表上构造评委侧汇总指标、周内相对指标以及淘汰标签，便于后续进行一致性检验以及后续建模。

1.2 清洗规则

(1) N/A 作为结构性缺失处理。 N/A 主要来自两类情形：当周不存在第4位评委（因此 judge4 缺失），或该赛季实际周数较少（超出该赛季的周列为 N/A）。因此保持为缺失值，并通过 num_judges 直接记录当周有效评委数量。

(2) 0 分视为淘汰后占位并转为缺失。 直接查看数据能够得出选手淘汰后剩余周会用 0 分占位。若将 0 当作真实得分，会在淘汰后影响周内排名与百分比。我们将 0 统一转为缺失（NaN），表示该周无参赛成绩。

1.3 重构流程

设赛季为 s，选手为 i，周次为 w，评委序号为 $k \in \{1,2,3,4\}$，评分记为 $x_{i,s,w,k}$。随后进行以下处理：

1. **识别周–评委评分列并宽转长：** 我们将 weekX_judgeY_score 列识别出来，解析得到 (w,k)，形成记录 $(s,i,w,k,x_{i,s,w,k})$。

2. **长转宽得到“选手–周”面板：** 将同一 (s,i,w) 下的评委得分透视为四列：

$$\texttt{judge1\_score},\ \texttt{judge2\_score},\ \texttt{judge3\_score},\ \texttt{judge4\_score}.$$

   若某赛季实际只有 3 名评委，则 judge4_score 保持缺失；若某个评委列在某些赛季从未出现，则显式补齐该列并填充缺失，保证列结构一致。

3. **静态信息回填：** 为避免静态字段存在缺失而导致面板记录被错误丢弃，面板构造时仅使用最小主键

$$(\texttt{season},\ \texttt{celebrity\_name},\ w),$$

   生成面板后再通过 (season, celebrity_name) 将静态信息合并回面板，从而最大程度保留全部选手记录。

1.4 评委侧特征构建

在面板数据中，对每个选手周 (i,w) 构造如下变量：

- 有效评委数：

$$n_{i,w} = \sum_{k=1}^{4} I(x_{i,w,k}\ \text{为有效数值}),$$

  输出为 num_judges；

- 评委总分：

$$J_{i,w} = \sum_{k=1}^{4} x_{i,w,k},$$

  输出为 judge_total；

- 周内评委排名 judge_rank：在同一赛季同一周内，仅对当周 $J_{i,w} > 0$ 的选手按 $J_{i,w}$ 降序排名（并列采用最小名次法）；

- 周内评委百分比：

$$p^{(J)}_{i,w} = \frac{J_{i,w}}{\sum_{j\in A_{s,w}} J_{j,w}}, \quad A_{s,w} = \{i \mid J_{i,w} > 0\},$$

  输出为 judge_percent。

1.5 淘汰标签：elimination_week 与 eliminated_this_week

后续模型需要检验“估计的粉丝票分配能否复现真实淘汰”，因此把淘汰信息结构化为标签：

- 若 `results` 包含 “Eliminated Week X”，直接提取 X 作为 `elimination_week`；

- 对 “2nd Place”、“3rd Place” 等不含周信息的情况，使用“最后一次有效参赛周”推断退出周。设赛季最大周

$$W_s = \max\{w \mid \exists i \text{ 使得 } J_{i,w} > 0\},$$

选手最后参赛周

$$w_i^{\text{last}} = \max\{w \mid J_{i,w} > 0\}.$$

若 $w_i^{\text{last}} < W_s$，则令 `elimination_week` 为 w_i^{last}；若 $w_i^{\text{last}} = W_s$，视为进入决赛周，`elimination_week` 留空。

- 对面板行 (i, w) 定义淘汰周指示变量

$$e_{i,w} = I(w = \texttt{elimination\_week}_i),$$

输出为 `eliminated_this_week`。

1.6 输出表与排序

清洗后输出 Excel 文件 `2026_MCM_C_clean_panel_week.xlsx`，包含两张面板表：`panel_by_week` 与 `panel_by_pairkey`。两表字段一致，仅排序不同：`panel_by_week` 按 (`season`, `week`, `judge_rank`, `celebrity_name`) 排序，便于周内比较；`panel_by_pairkey` 按 (`season`, `celebrity_name`, `week`) 排序，便于追踪单个选手的周序列。输出字段以 `celebrity_name`, `ballroom_partner`, `season` 开头，包含评委得分、周内排名/占比与淘汰标签等变量，方便直接观察以及后续一致性检验。

1.7 争议性样本识别（Table 3）

为刻画评委得分长期偏低但最终名次靠前的现象，我们基于面板表构造争议性样本表 `table3_controversy`。设赛季 s 第 w 周的有效参赛集合为

$$A_{s,w} = \{i \mid J_{i,s,w} > 0\}, \quad N_{s,w} = |A_{s,w}|.$$

定义 Bottom-K 指示变量（本文取 $K = 2$）：

$$b_{i,s,w} = I\big(\texttt{judge\_rank}_{i,s,w} \geq N_{s,w} - K + 1\big), \quad B_{i,s} = \sum_{w} b_{i,s,w}.$$

同时构造赛季级评委预期名次：先计算平均评委占比

$$\bar{p}^{(J)}_{i,s} = \frac{1}{|\Omega_{i,s}|} \sum_{w \in \Omega_{i,s}} p^{(J)}_{i,s,w},$$

再在赛季内按 $\bar{p}^{(J)}_{i,s}$ 降序排名得到 $\widehat{P}^{(J)}_{i,s}$。记实际最终名次为 $P_{i,s}$（即 `placement`），定义偏离

$$G_{i,s} = \widehat{P}^{(J)}_{i,s} - P_{i,s}.$$

我们仅在 $P_{i,s} \leq 5$ 的选手中筛选争议样本：若 $B_{i,s} \geq 5$ 或 $G_{i,s} \geq 3$ 则纳入表三，并在字段 `reason` 中记录触发原因。题面给出的典型案例（如第2季 Jerry Rice、第4季 Billy Ray Cyrus、第11季 Bristol Palin、第27季 Bobby Bones）强制纳入表三以便后续对照讨论。

2 Data Processing and Feature Construction

2.1 Data Structure and Processing Objectives

The raw dataset consists of two components. The first component is static information defined at the “contestant–season” level (e.g., `celebrity_name`, `ballroom_partner`, `celebrity_industry`, `celebrity_homestate`, `celebrity_homecountry_region`, `celebrity_age_during_season`, `results`, `placement`, etc.). The second component contains weekly judges’ scores, with column names in the form `weekX_judgeY_score`. Because we must first estimate each contestant's weekly fan votes before addressing the subsequent questions, we reorganize the data into a "contestant–week" panel (each row corresponds to one contestant's record in a given season and week). On this panel we construct aggregated judge-side metrics, within-week relative metrics, and elimination labels, which facilitates subsequent consistency checks and modeling.

2.2 Cleaning Rules

(1) Treat N/A as structural missingness. N/A mainly arises from two situations: there is no fourth judge in that week (hence `judge4` is missing), or the season has fewer actual weeks (so columns beyond the season's actual weeks are recorded as N/A). Therefore, we keep these entries as missing values and directly record the number of valid judges in that week using `num_judges`.

(2) Treat zero scores as post-elimination placeholders and convert them to missing. A direct inspection of the data shows that after a contestant is eliminated, the remaining weeks are filled with zeros as placeholders. If zeros are treated as genuine scores, they will distort within-week rankings and percentages after elimination. We therefore uniformly convert zeros to missing values (NaN), indicating that the contestant has no performance score in that week.

2.3 Reconstruction Procedure

Let the season be s, the contestant be i, the week be w, and the judge index be $k \in \{1, 2, 3, 4\}$. Denote the score by $x_{i,s,w,k}$. We then perform the following steps:

1. **Identify week–judge score columns and reshape from wide to long:** We identify columns of the form `weekX_judgeY_score`, parse (w, k), and form long-format records $(s, i, w, k, x_{i,s,w,k})$.

2. **Reshape from long to wide to obtain the "contestant–week" panel:** For each (s, i, w), we pivot judges' scores into four columns:

`judge1_score`, `judge2_score`, `judge3_score`, `judge4_score`.

If a season has only three judges in practice, `judge4_score` remains missing. If a judge column never appears in some seasons, we explicitly create that column and fill it with missing values to ensure a consistent column structure.

3. **Back-fill static information:** To avoid mistakenly dropping panel records due to missing values in static fields, we construct the panel using only the minimal primary key

(`season`, `celebrity_name`, w),

and then merge the static information back into the panel via (`season`, `celebrity_name`), thereby preserving contestant records to the greatest extent possible.

2.4 Construction of Judge-Side Features

In the panel data, for each contestant-week (i, w) we construct the following variables:

- Number of valid judges:

$$n_{i,w} = \sum_{k=1}^{4} I(x_{i,w,k} \text{ is a valid value}),$$

  output as `num_judges`;

- Total judges' score:

$$J_{i,w} = \sum_{k=1}^{4} x_{i,w,k},$$

  output as `judge_total`;

- Within-week judges' rank `judge_rank`: within the same season and week, we rank only contestants with $J_{i,w} > 0$ in descending order of $J_{i,w}$ (ties are handled by the minimum-rank method);

- Within-week judges' percentage:

$$p_{i,w}^{(J)} = \frac{J_{i,w}}{\sum_{j \in A_{s,w}} J_{j,w}}, \quad A_{s,w} = \{i \mid J_{i,w} > 0\},$$

  output as `judge_percent`.

2.5 Elimination Labels: elimination_week and eliminated_this_week

The subsequent model needs to test whether the estimated fan-vote allocation can reproduce the true eliminations; therefore, we structure elimination information as labels:

- If `results` contains "Eliminated Week X", we directly extract X as `elimination_week`;

- For cases such as "2nd Place" and "3rd Place" that do not contain week information, we infer the exit week using the "last week with a valid participation score." Let the maximum week of a season be

$$W_s = \max\{w \mid \exists\, i \text{ such that } J_{i,w} > 0\},$$

  and the contestant's last participation week be

$$w_i^{\text{last}} = \max\{w \mid J_{i,w} > 0\}.$$

  If $w_i^{\text{last}} < W_s$, we set `elimination_week` to w_i^{last}; if $w_i^{\text{last}} = W_s$, we treat the contestant as having reached the final week and leave `elimination_week` empty.

- For each panel row (i, w), define the indicator of the elimination week as

$$e_{i,w} = I(w = \texttt{elimination\_week}_i),$$

  output as `eliminated_this_week`.

2.6 Output Tables and Sorting Rules

After cleaning and feature construction, we export an Excel workbook `2026_MCM_C_clean_panel_we` containing two panel sheets with identical columns but different sorting orders. The sheet `panel_by_week` is sorted by

$$(\texttt{season},\ \texttt{week},\ \texttt{judge\_rank},\ \texttt{celebrity\_name}),$$

which facilitates within-week comparisons and elimination-consistency checks. The sheet `panel_by_pairkey` is sorted by

$$(\texttt{season},\ \texttt{celebrity\_name},\ \texttt{week}),$$

which is convenient for tracing each contestant' s week-by-week trajectory. Both sheets start with `celebrity_name`, `ballroom_partner`, and `season`, and include judge scores, within-week relative measures, and elimination labels for subsequent modeling.

2.7 Identifying Controversial Contestants (Table 3)

To quantify cases where a contestant receives persistently weak judge evaluations yet achieves a high final placement, we construct an auxiliary controversy table `table3_controversy` from the cleaned panel. Let the active set in season s and week w be

$$A_{s,w} = \{i \mid J_{i,s,w} > 0\}, \qquad N_{s,w} = |A_{s,w}|.$$

We define a Bottom-K indicator (with $K = 2$ in our implementation) by

$$b_{i,s,w} = I\big(\texttt{judge\_rank}_{i,s,w} \geq N_{s,w} - K + 1\big)\,, \qquad B_{i,s} = \sum_{w} b_{i,s,w},$$

where $B_{i,s}$ counts the number of weeks in which contestant i falls within the bottom-K by judge ranking. We further construct a season-level judge-implied expected placement by first computing the mean judge-share

$$\bar{p}^{(J)}_{i,s} = \frac{1}{|\Omega_{i,s}|} \sum_{w \in \Omega_{i,s}} p^{(J)}_{i,s,w},$$

ranking contestants within season s in descending order of $\bar{p}^{(J)}_{i,s}$ to obtain $\widehat{P}^{(J)}_{i,s}$, and comparing it to the realized final placement $P_{i,s}$ (i.e., `placement`) via

$$G_{i,s} = \widehat{P}^{(J)}_{i,s} - P_{i,s}.$$

We restrict attention to contestants finishing near the top ($P_{i,s} \leq 5$) and flag a contestant as controversial if $B_{i,s} \geq 5$ or $G_{i,s} \geq 3$. The field `reason` records which threshold(s) are exceeded. To align with the problem statement and support case-based discussion, the canonical examples (e.g., Jerry Rice in Season 2, Billy Ray Cyrus in Season 4, Bristol Palin in Season 11, and Bobby Bones in Season 27) are also included by construction.

3 模型一：粉丝票数估计、淘汰一致性与不确定性量化

3.1 目标与符号约定

考虑第 s 季节目，第 w 周仍在比赛中的组合集合记为 $\mathcal{A}_{s,w}$。对每个组合 $i \in \mathcal{A}_{s,w}$，定义其粉丝投票份额

$$x_{i,w} \in [0,1], \qquad \sum_{i \in \mathcal{A}_{s,w}} x_{i,w} = 1. \tag{1}$$

记该周总投票池规模为 $T_{s,w}$，则组合 i 的粉丝投票总数估计为

$$V_{i,w} = x_{i,w}\, T_{s,w}. \tag{2}$$

数据集中每周每位评委打分列形如 `weekX_judgeY_score`。记第 w 周评委总分为

$$J_{i,w} = \sum_{r=1}^{R_w} \mathrm{score}_{i,w,r}, \tag{3}$$

其中 R_w 为该周有效评委数量。集合 $\mathcal{A}_{s,w}$ 由“当周存在有效评委总分”确定。

3.2 投票池规模：收视人数到总票数的换算与缺失填补

每周收视数据提供 `total_viewers_millions`，因此

$$T_{s,w} = \mathrm{viewers}_{s,w} \times 10^6. \tag{4}$$

当 (s, w) 处收视数据缺失时，采用三级填补策略：同赛季按周次插值并前向/后向填充；同周次跨赛季选取季号距离最近的季作为替代。投票池规模仅用于份额到票数的换算。

3.3 粉丝投票份额点估计：库存项与表现项的可解释分解

份额估计首先构造当周吸票强度 $v_{i,w}$，再归一化得到先验份额 $\hat{x_{i,w}}$：

$$v_{i,w} = v_{i,w}^{\mathrm{star}} + v_{i,w}^{\mathrm{pro}} + v_{i,w}^{\mathrm{perf}} + \varepsilon, \qquad \hat{x_{i,w}} = \frac{v_{i,w}}{\sum_{j\in\mathcal{A}_{s,w}} v_{j,w}}, \tag{5}$$

其中 $\varepsilon > 0$ 为数值稳定项。三部分含义如下。

3.3.1 明星/舞者粉丝库存与赛季初始收缩混合

对明星与职业舞者分别维护随周更新的粉丝库存（log 空间）：

$$\ell_{i,w}^{\mathrm{star}} = \log F_{i,w}^{\mathrm{star}}, \qquad \ell_{i,w}^{\mathrm{pro}} = \log F_{i,w}^{\mathrm{pro}}. \tag{6}$$

赛季初始值采用“跨赛季全局基线 + 本赛季猜测”的收缩混合：

$$\ell_{i,0}^{\mathrm{star}} = \rho_{\mathrm{star}}\, g_{\mathrm{name}(i)}^{\mathrm{star}} + (1-\rho_{\mathrm{star}})\, \log(\mathrm{star\_guess}_i + 1), \tag{7}$$

$$\ell_{i,0}^{\mathrm{pro}} = \rho_{\mathrm{pro}}\, g_{\mathrm{pro}(i)}^{\mathrm{pro}} + (1-\rho_{\mathrm{pro}})\, \log(\mathrm{pro\_guess}_i + 1), \tag{8}$$

其中 g^{star} 与 g^{pro} 为跨赛季共享的基线库存（见第 3.6 小节），$\rho_{\mathrm{star}}, \rho_{\mathrm{pro}} \in [0, 1]$ 为收缩系数。$\mathrm{star\_guess}_i$ 由赛季第 1 周评委总分经指数加权分配得到，用于体现“开局表现强者更高的初始关注度”。

由库存得到两类固定票强度：

$$v_{i,w}^{\mathrm{star}} = \exp(\ell_{i,w}^{\mathrm{star}}), \qquad v_{i,w}^{\mathrm{pro}} = \alpha_{\mathrm{pro}} \exp(\ell_{i,w}^{\mathrm{pro}}), \tag{9}$$

其中 α_{pro} 为舞者库存到固定票强度的换算系数。

3.3.2 评委表现特征与当周胶着度

当周评委总分向量记为 $J_w = \{J_{i,w}\}_{i\in\mathcal{A}_{s,w}}$。定义评委百分比为

$$p_{i,w} = \frac{J_{i,w}}{\sum_{j\in\mathcal{A}_{s,w}} J_{j,w}}. \tag{10}$$

定义评委名次 $\mathrm{rank}^J_{i,w}$（1 为最好），并将其映射到 $[0,1]$ 的强度分数：

$$q_{i,w} = \frac{n_w - \mathrm{rank}^J_{i,w} + 1}{n_w}, \qquad n_w = |\mathcal{A}_{s,w}|. \tag{11}$$

为刻画当周竞争胶着程度，先定义相对离散度

$$\mathrm{spread}_w = \frac{\max(J_w) - \min(J_w)}{\overline{J_w} + \varepsilon}, \tag{12}$$

再经 sigmoid 映射得到

$$\mathrm{tight}_w = \sigma\left(k_{\mathrm{tight}}(\tau_{\mathrm{tight}} - \mathrm{spread}_w)\right), \tag{13}$$

其中 k_{tight} 与 τ_{tight} 为超参数；tight_w 越大表示当周越胶着。

3.3.3 当周表现票

当周表现票以 $q_{i,w}$ 为核心，并显式建模“优秀加成”与“危险区动员”。设危险区为按评委名次选取的 bottom-K 集合，则

$$v^{\mathrm{perf}}_{i,w} = \beta_{\mathrm{perf}}\left(q_{i,w}^{a_{\mathrm{exc}}} + \omega_{\mathrm{dang}} \cdot \mathrm{tight}_w \cdot \mathbf{1}\{i \in \text{bottom-}K\} \cdot (1 - q_{i,w})^{a_{\mathrm{dang}}}\right), \tag{14}$$

其中 β_{perf} 为表现票尺度，a_{exc} 控制优秀加成非线性，ω_{dang} 与 a_{dang} 控制危险区动员强度与形状。

3.4 淘汰一致性约束

点估计 $\hat{x}_{i,w}$ 提供“先验倾向”，为符合题目给定的淘汰机制，进一步对每周份额进行校正得到 $x_{i,w}$。节目规则按赛季分段：

3.4.1 Seasons 1–2：rank + rank

评委与粉丝均按排名计分。由份额 x_w 得到粉丝名次 $\mathrm{rank}^F_{i,w}$（份额越大名次越靠前）。综合劣势分数定义为

$$C_{i,w} = w_J\,\mathrm{rank}^J_{i,w} + w_F\,\mathrm{rank}^F_{i,w}. \tag{15}$$

设当周淘汰者集合为 E_w（大小 k_w），一致性约束为

$$E_w \subseteq \text{Worst-}k_w(C_w). \tag{16}$$

由于排名函数离散不可微，校正采用最小扰动迭代：以 $\hat{x_w}$ 归一化结果为初值，若淘汰者不在 Worst-k_w，则以小步长削减淘汰者份额并将削减质量按当前分布回填其余选手，直至满足约束或达到迭代上限。该过程在保持单纯形约束的同时尽量贴近先验份额。

3.4.2 Seasons 3–27：percent + percent

评委以百分比计分，粉丝以份额计分。综合分数定义为

$$C_{i,w} = w_J\,p_{i,w} + w_F\,x_{i,w}. \tag{17}$$

对淘汰者 e 与幸存者 j，要求 $C_{e,w} \le C_{j,w}$，等价于

$$w_F(x_{e,w} - x_{j,w}) \le w_J(p_{j,w} - p_{e,w}). \tag{18}$$

为保证可行性，引入 slack 变量 $z_k \ge 0$：

$$w_F(x_{e,w} - x_{j,w}) - z_k \le w_J(p_{j,w} - p_{e,w}), \qquad z_k \ge 0. \tag{19}$$

校正通过凸二次规划实现：

$$\min_{x\in\Delta,\ z\ge 0} \quad \lambda_{\text{prior}}\|x - \hat{x}\|_2^2 + \lambda_{\text{slack}}\sum_k z_k \tag{20}$$

$$\text{s.t.} \quad w_F(x_{e,w} - x_{j,w}) - z_k \le w_J(p_{j,w} - p_{e,w}),\ \forall j \in \mathcal{A}_{s,w} \setminus \{e\}, \tag{21}$$

其中 Δ 为单纯形。λ_{prior} 控制对先验份额的贴近程度，λ_{slack} 控制对淘汰逻辑的约束强度。周级 slack 总量用于度量“为实现淘汰一致性所需的最小违约”。

3.4.3 Seasons 28–34：rank + rank + bottom-2 lock

该阶段机制为先锁定 bottom-2，再由评委在直播中决定淘汰。校正仅强制淘汰者属于 bottom-2 集合。综合劣势分数仍取

$$C_{i,w} = w_J \,\mathrm{rank}^J_{i,w} + w_F \,\mathrm{rank}^F_{i,w}, \tag{22}$$

并要求

$$E_w \subseteq \mathrm{Worst\text{-}2}(C_w). \tag{23}$$

实现方式沿用 rank 场景下的最小扰动迭代策略。

3.5 粉丝库存的周更新：表现驱动增益/流失与衰减

在获得校正份额后，更新明星与舞者库存以影响下一周的固定票。定义

$$\mathrm{exc}_{i,w} = q_{i,w}^{\gamma_{\mathrm{gain}}}, \tag{24}$$

$$\mathrm{dang}_{i,w} = \mathrm{tight}_w \cdot \mathbf{1}\{i \in \mathrm{bottom}\text{-}K\} \cdot (1 - q_{i,w})^{\gamma_{\mathrm{dang}}}, \tag{25}$$

$$\mathrm{loss}_{i,w} = (1 - q_{i,w})^{\gamma_{\mathrm{loss}}}, \tag{26}$$

并令

$$\mathrm{gain}_{i,w} = \mathrm{exc}_{i,w} + \omega_{\mathrm{dang}} \cdot \mathrm{dang}_{i,w}. \tag{27}$$

则 log 库存按周递推：

$$\ell^{\mathrm{star}}_{i,w+1} = \ell^{\mathrm{star}}_{i,w} + \eta^{\mathrm{star}}_{+}\,\mathrm{gain}_{i,w} - \eta^{\mathrm{star}}_{-}\,\mathrm{loss}_{i,w} - \delta^{\mathrm{star}}, \tag{28}$$

$$\ell^{\mathrm{pro}}_{i,w+1} = \ell^{\mathrm{pro}}_{i,w} + \eta^{\mathrm{pro}}_{+}\,\mathrm{gain}_{i,w} - \eta^{\mathrm{pro}}_{-}\,\mathrm{loss}_{i,w} - \delta^{\mathrm{pro}}. \tag{29}$$

该更新机制体现“高表现带来人气积累，低表现带来流失，自然衰减长期存在；危险区在胶着周触发额外动员”的动态规律。

3.6 跨赛季外循环：全局基线的经验更新与收缩稳定性

模型在所有赛季上执行多轮外循环。每轮外循环完成全部赛季求解后，收集赛季初始 log 库存作为观测量，对同名明星与同名舞者分别取均值，并对全局基线

做指数滑动更新：

$$g \leftarrow (1-\rho)\, g + \rho \cdot \overline{\ell_0}, \tag{30}$$

其中 $\rho \in (0,1)$ 为平滑系数，$\overline{\ell_0}$ 为该主体在当前外循环中的初始 log 库存均值。最后一轮外循环输出完整结果，从而得到更稳定的跨赛季基线与更一致的份额估计。

3.7 淘汰一致性评估指标

对赛季 s 的每个淘汰周 w，定义一致性指示变量

$$\mathbb{I}_{s,w} = \begin{cases} 1, & \text{真实淘汰者满足当季规则对应的 bottom 集合约束}, \\ 0, & \text{否则}. \end{cases} \tag{31}$$

赛季一致性率定义为

$$\text{ConsistencyRate}_s = \frac{1}{|\mathcal{W}_s|} \sum_{w \in \mathcal{W}_s} \mathbb{I}_{s,w}, \tag{32}$$

其中 $\mathcal{W}_s$ 为赛季 s 所有发生淘汰的周集合。对 percent 机制赛季，另输出 slack 总量用于衡量约束冲突程度。

3.8 不确定性量化

3.8.1 可行域区间代理

在 percent + percent 赛季，约束可写为线性不等式且 x 位于单纯形上，可在同一可行域内对每个分量分别求

$$x_{i,w}^{\min} = \min x_{i,w}, \qquad x_{i,w}^{\max} = \max x_{i,w}, \tag{33}$$

并以

$$U_{i,w}^{x} = x_{i,w}^{\max} - x_{i,w}^{\min} \tag{34}$$

作为份额不确定性的区间代理。rank 场景由离散排序诱导可行域，线性区间代理不稳定，因此该代理主要用于 percent 机制赛季。

3.8.2 贝叶斯一致性分析（后验命中概率与可信区间）

为给出概率层面的确定性度量，并显式量化每周/每位参赛者估计值的不确定性，对每个发生淘汰的周构造份额向量 $x_w = (x_{i,w})_{i\in\mathcal{A}_{s,w}}$ 的后验分布。以点估计份额 $x_w^{(0)}$ 为中心，引入 Dirichlet 先验

$$x_w \sim \mathrm{Dirichlet}(\alpha_w), \qquad \alpha_w = \kappa_w x_w^{(0)}, \tag{35}$$

其中浓度参数 $\kappa_w > 0$ 控制先验的集中程度：κ_w 越大表示越信任点估计，κ_w 越小表示允许围绕 $x_w^{(0)}$ 更大幅度波动。为在不同周之间共享信息并避免逐周独立估计造成的数值不稳定，对 κ_w 引入分层先验

$$\log \kappa_w \sim \mathcal{N}(\mu, \tau^2), \tag{36}$$

并采用经验贝叶斯外循环迭代更新 (μ, τ)：在每轮中先对所有周分别求得 $\log \kappa_w$ 的 MAP 估计，再以其样本均值与标准差更新 (μ, τ)，从而得到跨周稳定的一致性先验强度。

淘汰事件通过成对比较构建似然。记当周合成分数为 $C_i(x)$（按赛季规则取 percent 或 rank 形式）。对淘汰者 e 与任一幸存者 j，用 logistic 形式刻画“淘汰者更差”的概率：

$$\Pr(e \text{ 劣于 } j \mid x) \approx \sigma\left(\frac{C_j(x) - C_e(x)}{\sigma_w}\right), \qquad \sigma(z) = \frac{1}{1+e^{-z}}. \tag{37}$$

其中 σ_w 为周级噪声尺度，刻画淘汰机制的随机性与不可观测因素强度；该尺度由当周竞争胶着度 tight_w 调节：

$$\sigma_w = \sigma_{\mathrm{base}}\big(1 + \sigma_{\mathrm{tight}} \cdot \mathrm{tight}_w\big). \tag{38}$$

当周越胶着（tight_w 越大），σ_w 越大，表示淘汰边界更不清晰，后验不确定性相应增加。对于 Season 28+ 的 bottom-2 lock 机制，似然不再要求淘汰者为唯一最差者，而是以“淘汰者属于 bottom-2 集合”的概率构造对应的可微混合形式似然，以匹配“锁定最后两名后再由评委二选一”的规则结构。

后验推断采用两阶段计算。首先对每周后验目标进行 MAP 优化得到 x_w^{MAP} 与 κ_w^{MAP}。为在单纯形约束下进行无约束优化与二阶近似，将 x 映射到 Aitchison 对

数比（ALR）坐标：

$$y_k = \log \frac{x_k}{x_n}, \qquad k = 1, \ldots, n-1, \quad n = |\mathcal{A}_{s,w}|. \tag{39}$$

其次在 y^{MAP} 处进行 Laplace 近似：计算负对数后验的 Hessian 并取协方差近似 $\Sigma \simeq (H + \epsilon I)^{-1}$（必要时加入对角 jitter 与正定投影以保证数值稳定），在近似高斯下采样 $y^{(m)} \sim \mathcal{N}(y^{\text{MAP}}, \Sigma)$，再映射回 $x^{(m)}$。由后验样本得到份额可信区间与宽度指标

$$[x_{i,w}^{\text{low}}, x_{i,w}^{\text{high}}], \qquad U_{i,w}^{x} = x_{i,w}^{\text{high}} - x_{i,w}^{\text{low}}. \tag{40}$$

进一步由 $V_{i,w} = x_{i,w} T_{s,w}$ 将份额样本转换为票数样本 $V_{i,w}^{(m)} = x_{i,w}^{(m)} T_{s,w}$，得到票数可信区间

$$[V_{i,w}^{\text{low}}, V_{i,w}^{\text{high}}], \qquad U_{i,w}^{V} = V_{i,w}^{\text{high}} - V_{i,w}^{\text{low}}. \tag{41}$$

由此可直接检验不确定性在不同参赛者与不同周之间是否一致：例如比较同一周内 $\{U_{i,w}^{x}\}$ 的分布，或比较同一参赛者跨周 $\{U_{i,w}^{x}\}$ 的统计量；亦可使用相对宽度 $U_{i,w}^{x}/x_{i,w}^{\text{MAP}}$ 与 $U_{i,w}^{V}/V_{i,w}^{\text{MAP}}$ 排除规模效应。

同时定义后验淘汰一致性概率（后验命中概率）

$$\text{PosteriorHitProb}_{s,w} = \Pr(\text{当季淘汰规则在后验样本下成立} \mid \text{data}) \approx \frac{1}{M} \sum_{m=1}^{M} \mathbf{1}\big\{e \in \text{BottomSet}\big(C(x^{(m)} \tag{42}$$

其中 $\text{BottomSet}(\cdot)$ 按赛季规则分别取 bottom-m（percent）、worst-k（rank）或 bottom-2（lock）。此外输出周级对比强度指标用于解释命中概率与区间宽度的变化来源：其一为 MAP 点处的平均成对概率（将上述 logistic 成对概率对幸存者取平均）；其二为边际指标（淘汰者与淘汰阈值之间的合成分数差），用于刻画淘汰边界的分离度与置信程度。

3.9 遗传算法调参：以贝叶斯一致性为目标的黑盒优化

为提升后验命中概率并抑制不确定性宽度，对估计器参数向量 θ 进行遗传算法优化。每个个体对应一组 θ，其评估流程为：运行估计器生成点估计表，运行贝

叶斯一致性分析输出周级指标，再汇总得到适应度函数

$$\text{Fitness}(\theta) = a \cdot \overline{\text{PosteriorHitProb}} + b \cdot \overline{\text{AvgPairProb}} + c \cdot \overline{\text{Margin}} - d \cdot \overline{U^x} - e \cdot \text{FailRate}, \tag{43}$$

其中 $\bar{\cdot}$ 表示跨周/跨赛季平均或加权平均，FailRate 为数值优化失败周比例。遗传算法采用标准选择–交叉–变异–精英保留框架，并通过多进程并行加速个体评估；断点续跑、播种策略（seed-once）、搜索空间定义、日志与产物组织等实现细节见配套 README 文档。

4 Professional Dancer and Celebrity-Feature Effects on Judges and Fans

4.1 Data and outcomes

Let i index pairs (celebrity–professional), (s, w) index season and week, and let $\mathcal{A}_{s,w}$ be the set of active pairs in week w of season s.

Judges. Let $J_{i,s,w} \in (0, 1)$ be the within-week judge share (`judge_percent`). Define

$$\tilde{J}_{i,s,w} = \text{logit}(J_{i,s,w}), \qquad y^{(J)}_{i,s,w} = \tilde{J}_{i,s,w} - \frac{1}{|\mathcal{A}_{s,w}|} \sum_{k \in \mathcal{A}_{s,w}} \tilde{J}_{k,s,w}. \tag{44}$$

Thus $y^{(J)}$ measures relative judge advantage within each (s, w), removing week-level shocks.

Fans. From dwts_fan_vote_estimates_components.csv, we use the estimated vote-share components: $P^{(\text{tot})}$, $P^{(\star)}$ (celebrity fan-base), $P^{(\text{pro})}$ (pro-fixed), and $P^{(\text{perf})}$ (performance-induced). For $o \in \{\text{tot}, \star, \text{pro}\}$ we define

$$\tilde{P}^{(o)}_{i,s,w} = \text{logit}\Big(P^{(o)}_{i,s,w}\Big), \qquad y^{(o)}_{i,s,w} = \tilde{P}^{(o)}_{i,s,w} - \frac{1}{|\mathcal{A}_{s,w}|} \sum_{k \in \mathcal{A}_{s,w}} \tilde{P}^{(o)}_{k,s,w}. \tag{45}$$

For the performance component we use a log transform (to accommodate very small shares):

$$\tilde{P}^{(\mathrm{perf})}_{i,s,w} = \log\Big(P^{(\mathrm{perf})}_{i,s,w} + \varepsilon\Big), \qquad y^{(\mathrm{perf})}_{i,s,w} = \tilde{P}^{(\mathrm{perf})}_{i,s,w} - \frac{1}{|\mathcal{A}_{s,w}|}\sum_{k\in\mathcal{A}_{s,w}} \tilde{P}^{(\mathrm{perf})}_{k,s,w}. \tag{46}$$

4.2 Explanatory factors

We include all celebrity features present in the panel: `celebrity_age_during_season`, `celebrity_industry`, `celebrity_homestate`, and `celebrity_homecountry_region`. Age is standardized:

$$\mathrm{age\_z}_{i,s} = \frac{\mathrm{age}_{i,s} - \overline{\mathrm{age}}}{\mathrm{sd}(\mathrm{age})}. \tag{47}$$

Industries and states are pooled by minimum frequency, and regions are top-N pooled, yielding grouped categories ind_i, state_i, reg_i. The professional dancer identity is pro_i (`ballroom_partner`).

4.3 Model specification

For each outcome $o \in \{J, \mathrm{tot}, \star, \mathrm{pro}, \mathrm{perf}\}$ we fit:

$$y^{(o)}_{i,s,w} = \beta^{(o)}_{\mathrm{age}}\,\mathrm{age\_z}_{i,s} + \alpha^{(o)}_{\mathrm{ind}(i)} + \alpha^{(o)}_{\mathrm{state}(i)} + \alpha^{(o)}_{\mathrm{reg}(i)} + \gamma^{(o)}_{\mathrm{pro}(i)} + \varepsilon^{(o)}_{i,s,w}, \tag{48}$$

with category fixed effects α and professional fixed effects γ. Standard errors are clustered at the season–celebrity level (same celebrity across weeks).

4.4 Evaluation metrics

(i) Incremental explanatory power of professionals. Define $R^2_{\mathrm{no\_pro}}(o)$ from the model without $\gamma^{(o)}_{\mathrm{pro}(i)}$, and $R^2_{\mathrm{full}}(o)$ from the full model. Then

$$\Delta R^2_{\mathrm{pro}}(o) = R^2_{\mathrm{full}}(o) - R^2_{\mathrm{no\_pro}}(o). \tag{49}$$

(ii) Factor importance via drop-one ΔR^2. For factor $f \in \{\text{age}, \text{industry}, \text{state}, \text{region}, \text{pro}\}$, let $R^2_{-f}(o)$ be the R^2 from the model dropping only factor f, fit on the same complete-case sample as the full model. Define

$$\Delta R^2_{\text{drop}}(o, f) = R^2_{\text{full}}(o) - R^2_{-f}(o). \tag{50}$$

(iii) Cross-outcome alignment (same-way test). For a categorical factor f with category effects $\theta^{(o)}_{f,c}$ (centered by category frequency), define

$$\text{Align}_f(o) = \text{Corr}\left(\theta^{(J)}_{f,\cdot},\, \theta^{(o)}_{f,\cdot}\right). \tag{51}$$

(iv) Age effect as a per-10-year multiplier. Let σ_{age} be the age standard deviation (years). The multiplicative change per +10 years is

$$\text{Mult}^{(o)}_{\text{age}}(+10) = \exp\left(\beta^{(o)}_{\text{age}} \cdot \frac{10}{\sigma_{\text{age}}}\right). \tag{52}$$

4.5 Results

4.5.1 Incremental explanatory power of professionals beyond celebrity features

Table 1 reports R^2 from the model without professional fixed effects and from the full model, fitted on the same complete-case sample ($n = 2402$). Define

$$\Delta R^2_{\text{pro}}(o) = R^2_{\text{full}}(o) - R^2_{\text{no\_pro}}(o). \tag{53}$$

Table 1 shows that, after controlling for age, industry, state, and region, professional identity still increases explanatory power by 0.052–0.099 absolute R^2 across outcomes, with the largest gains on Judges and on FansTotal/FansStar.

4.5.2 Marginal contribution (drop-one ΔR^2)

For each factor $f \in \{\text{pro}, \text{industry}, \text{state}, \text{region}, \text{age}\}$, define

$$\Delta R^2_{\text{drop}}(o, f) = R^2_{\text{full}}(o) - R^2_{-f}(o), \tag{54}$$

表 1: R^2 comparison and incremental explanatory power of professionals (all celebrity features controlled)

Outcome o	$R^2_{\text{no\_pro}}(o)$	$R^2_{\text{full}}(o)$	$\Delta R^2_{\text{pro}}(o)$
Judges	0.2584	0.3577	0.0994
FansTotal	0.1565	0.2487	0.0922
FansStar	0.1624	0.2547	0.0923
FansPro	0.0655	0.1246	0.0591
FansPerf	0.1624	0.2143	0.0520

where $R^2_{-f}(o)$ is obtained by dropping only f and refitting on the same sample as the full model.

表 2: Factor importance via drop-one ΔR^2 (larger values indicate larger contributions)

Outcome o	$\Delta R^2_{\text{drop}}(o, \text{pro})$	$\Delta R^2_{\text{drop}}(o, \text{industry})$	$\Delta R^2_{\text{drop}}(o, \text{state})$	$\Delta R^2_{\text{drop}}(o, \text{region})$	$\Delta R^2_{\text{drop}}(o, \text{age})$
Judges	0.0994	0.0320	0.0281	-0.0016	0.0885
FansTotal	0.0922	0.0125	0.0330	-0.0001	0.0393
FansStar	0.0923	0.0135	0.0333	-0.0001	0.0412
FansPro	0.0591	0.0065	0.0144	-0.0000	0.0180
FansPerf	0.0520	0.0248	0.0264	-0.0005	0.0585

Table 2 implies: professionals are first-order drivers for every outcome, contributing about 0.052–0.099 drop-one R^2; age is the strongest celebrity feature, nearly matching professionals for Judges (0.0885) and remaining large for FansPerf (0.0585); state and industry provide moderate explanatory power; (iv) region contributes essentially zero. The tiny negative values for region are numerical/sample-boundary artifacts and should be interpreted as $\Delta R^2_{\text{drop}}(o, \text{region}) \approx 0$.

4.5.3 Mechanism alignment

For each categorical factor $f \in \{\text{pro}, \text{industry}, \text{state}\}$, let $\theta^{(o)}_{f,\cdot}$ denote the centered category-effect vector under outcome o. Define

$$\mathrm{Align}_f(o) = \mathrm{Corr}\left(\theta^{(J)}_{f,\cdot},\, \theta^{(o)}_{f,\cdot}\right). \tag{55}$$

表 3: Alignment between judges and fan outcomes (correlation of category effects)

Factor f	$\mathrm{Align}_f(\mathrm{tot})$	$\mathrm{Align}_f(\star)$	$\mathrm{Align}_f(\mathrm{pro})$	$\mathrm{Align}_f(\mathrm{perf})$
pro	0.7546	0.7626	0.5292	0.8933
industry	0.7457	0.7555	0.6036	0.9650
state	0.7181	0.7295	0.4907	0.6901

Table 3 shows moderate alignment on FansTotal/FansStar (≈ 0.75), maximal alignment on the performance component (pro: 0.8933, industry: 0.9650), and minimal alignment on the pro-fixed component (pro: 0.5292, state: 0.4907); thus effects are not transmitted through identical channels across judges and fans.

4.5.4 Age effect across outcomes

Let age enter standardized with standard deviation $\sigma_{\text{age}} = 12.8691$ years. The per-10-year multiplier is

$$\mathrm{Mult}^{(o)}_{\text{age}}(+10) = \exp\left(\beta^{(o)}_{\text{age}} \cdot \frac{10}{\sigma_{\text{age}}}\right). \tag{56}$$

Table 4 implies a mild judge-side penalty (Judges: 0.954 per +10y, i.e., $\approx 4.6\%$ decrease) but a strong fan-side penalty (FansTotal/FansStar: ≈ 0.66, i.e., ≈ 33–34% decrease), strongest on FansPerf (0.637), indicating age primarily operates through fan-side channels.

表 4: Age effects: coefficients and per +10-year multipliers ($\sigma_{\text{age}} = 12.8691$ years)

Outcome o	$\beta^{(o)}_{\text{age}}$ (on age_z)	$\text{Mult}^{(o)}_{\text{age}}(+10)$
Judges	-0.060852	0.953815
FansTotal	-0.519122	0.668055
FansStar	-0.535738	0.659485
FansPro	-0.214542	0.846444
FansPerf	-0.580932	0.636727

4.6 Answer to the question

Professionals and celebrity features jointly shape performance: professionals add $\Delta R^2_{\text{pro}}(o) \in [0.052, 0.099]$ beyond celebrity features (Table 1) and are among the most important drivers across outcomes (Table 2); among celebrity features, age is strongest and affects fans far more than judges (Table 4), while industry and state are moderate contributors. Mechanistically, effects are not identical across judges and fans: alignment peaks on the performance component and is weakest on the pro-fixed component (Table 3), implying judges track performance-driven variation most consistently whereas fan totals also reflect fixed/support-base channels.

5 职业舞者效应与名人特征对评委与粉丝的影响

5.1 数据与结果变量

用 i 表示组合（名人–职业舞者），(s, w) 表示赛季与周次，$\mathcal{A}_{s,w}$ 表示赛季 s 第 w 周仍在比赛的组合集合。我们在合并后的面板数据上进行分析，完整样本量为 $n = 2402$（complete cases）。

评委端。 令 $J_{i,s,w} \in (0, 1)$ 为当周评委份额（`judge_percent`）。定义

$$\tilde{J}_{i,s,w} = \text{logit}(J_{i,s,w}), \qquad y^{(J)}_{i,s,w} = \tilde{J}_{i,s,w} - \frac{1}{|\mathcal{A}_{s,w}|} \sum_{k \in \mathcal{A}_{s,w}} \tilde{J}_{k,s,w}. \tag{57}$$

因此 $y^{(J)}$ 是同一 (s,w) 内的相对评委优势，已剔除周层公共冲击。

粉丝端（分量）。 来自 `dwts_fan_vote_estimates_components.csv`，我们使用估计的投票份额分量：总份额 $P^{(\text{tot})}$、名人粉丝底盘 $P^{(\star)}$（celebrity fan-base）、职业舞者固定盘 $P^{(\text{pro})}$（pro-fixed）、以及表现诱导盘 $P^{(\text{perf})}$（performance-induced）。对 $o\in\{\text{tot},\star,\text{pro}\}$ 定义

$$\tilde{P}^{(o)}_{i,s,w}=\text{logit}\Big(P^{(o)}_{i,s,w}\Big),\qquad y^{(o)}_{i,s,w}=\tilde{P}^{(o)}_{i,s,w}-\frac{1}{|\mathcal{A}_{s,w}|}\sum_{k\in\mathcal{A}_{s,w}}\tilde{P}^{(o)}_{k,s,w}. \tag{58}$$

对表现分量由于份额极小，采用对数变换（默认）：

$$\tilde{P}^{(\text{perf})}_{i,s,w}=\log\Big(P^{(\text{perf})}_{i,s,w}+\varepsilon\Big),\qquad y^{(\text{perf})}_{i,s,w}=\tilde{P}^{(\text{perf})}_{i,s,w}-\frac{1}{|\mathcal{A}_{s,w}|}\sum_{k\in\mathcal{A}_{s,w}}\tilde{P}^{(\text{perf})}_{k,s,w}. \tag{59}$$

5.2 解释变量（所有可得名人特征）

我们纳入面板中**所有**可得的名人特征：赛季年龄（`celebrity_age_during_season`）、行业（`celebrity_industry`）、家乡州（`celebrity_homestate`）、以及国家/地区大区（`celebrity_homecountry_region`）。年龄做标准化：

$$\text{age\_z}_{i,s}=\frac{\text{age}_{i,s}-\overline{\text{age}}}{\text{sd}(\text{age})}. \tag{60}$$

行业与州对低频类别按最小频数阈值合并，地区按 top-N 保留其余合并，得到分组类别 $\text{ind}_i,\text{state}_i,\text{reg}_i$。职业舞者身份记为 pro_i（`ballroom_partner`）。

5.3 模型设定

对每个结果 $o\in\{J,\text{tot},\star,\text{pro},\text{perf}\}$ 拟合：

$$y^{(o)}_{i,s,w}=\beta^{(o)}_{\text{age}}\,\text{age\_z}_{i,s}+\alpha^{(o)}_{\text{ind}(i)}+\alpha^{(o)}_{\text{state}(i)}+\alpha^{(o)}_{\text{reg}(i)}+\gamma^{(o)}_{\text{pro}(i)}+\varepsilon^{(o)}_{i,s,w}, \tag{61}$$

其中 α 为类别固定效应，γ 为职业舞者固定效应。标准误采用 season–celebrity 维度聚类（同一名人在多周重复观测）。

5.4 评价指标

(i) 职业舞者的增量解释力。 令 $R^2_{\text{no\_pro}}(o)$ 为删除 $\gamma^{(o)}_{\text{pro}(i)}$ 的模型 R^2， $R^2_{\text{full}}(o)$ 为完整模型 R^2，定义

$$\Delta R^2_{\text{pro}}(o) = R^2_{\text{full}}(o) - R^2_{\text{no\_pro}}(o). \tag{62}$$

(ii) 因素重要性：drop-one ΔR^2。 对因素 $f \in \{\text{age}, \text{industry}, \text{state}, \text{region}, \text{pro}\}$，令 $R^2_{-f}(o)$ 为仅删除因素 f 的模型 R^2（在与完整模型相同的 complete-case 样本上拟合），定义

$$\Delta R^2_{\text{drop}}(o, f) = R^2_{\text{full}}(o) - R^2_{-f}(o). \tag{63}$$

(iii) 跨结果对齐度（是否“同一种方式”影响）。 对类别因素 f，其类别效应（按类别频数中心化后）记为 $\theta^{(o)}_{f,c}$，定义

$$\text{Align}_f(o) = \text{Corr}\left(\theta^{(J)}_{f,\cdot}, \theta^{(o)}_{f,\cdot}\right). \tag{64}$$

(iv) 年龄效应：每增加 10 岁的乘子。 令 σ_{age} 为年龄标准差（年），则每增加 10 岁的乘子为

$$\text{Mult}^{(o)}_{\text{age}}(+10) = \exp\left(\beta^{(o)}_{\text{age}} \cdot \frac{10}{\sigma_{\text{age}}}\right). \tag{65}$$

5.5 结果

5.5.1 职业舞者在控制全部名人特征后的增量解释力

表 1 给出在同一 complete-case 样本（$n = 2402$）上，删除职业舞者固定效应与完整模型的 R^2 对比。定义

$$\Delta R^2_{\text{pro}}(o) = R^2_{\text{full}}(o) - R^2_{\text{no\_pro}}(o). \tag{66}$$

表 1 表明：在控制年龄、行业、州、地区后，职业舞者身份仍能在所有结果上带来 0.052–0.099 的绝对 R^2 提升，其中对评委端以及 FansTotal/FansStar 的提升最大。

5.5.2 全部名人特征与职业舞者的相对重要性（drop-one ΔR^2）

对每个因素 $f \in \{\text{pro}, \text{industry}, \text{state}, \text{region}, \text{age}\}$，定义

$$\Delta R^2_{\text{drop}}(o, f) = R^2_{\text{full}}(o) - R^2_{-f}(o), \tag{67}$$

其中 $R^2_{-f}(o)$ 为仅删除因素 f 后在相同样本上重新拟合得到的 R^2。

表 2 表明：（i）职业舞者对所有结果都是一阶驱动因素，drop-one 贡献约为 0.052–0.099；（ii）年龄是最强名人特征，在 Judges 上几乎与职业舞者同量级（0.0885），在 FansPerf 上也很大（0.0585）；（iii）州与行业具有中等解释力；（iv）地区贡献基本为零。地区出现的极小负值来自数值/样本边界效应，应解释为 $\Delta R^2_{\text{drop}}(o, \text{region}) \approx 0$。

5.5.3 是否以相同方式影响评委与粉丝：跨结果对齐度

对每个类别因素 f（职业舞者、行业、州），令 $\theta^{(o)}_{f,\cdot}$ 为结果 o 下的中心化类别效应向量，定义

$$\text{Align}_f(o) = \text{Corr}\left(\theta^{(J)}_{f,\cdot}, \theta^{(o)}_{f,\cdot}\right). \tag{68}$$

表 3 直接回答“是否相同方式影响评委与粉丝”：职业舞者与行业在 FansTotal/FansStar 上具有中等对齐（≈ 0.75），在表现分量上对齐最高（pro: 0.8933；industry: 0.9650），在职业舞者固定盘分量上对齐最低（pro: 0.5292；state: 0.4907）。因此，同一类别并不通过完全相同通道影响评委与粉丝；最高一致性集中在表现驱动分量上。

5.5.4 年龄效应跨结果对比（每 +10 岁乘子）

年龄以标准化变量进入模型。令年龄标准差为 σ_{age}（年），定义每增加 10 岁的乘子：

$$\text{Mult}^{(o)}_{\text{age}}(+10) = \exp\left(\beta^{(o)}_{\text{age}} \cdot \frac{10}{\sigma_{\text{age}}}\right). \tag{69}$$

表 4 表明：年龄对粉丝端的惩罚远强于评委端。+10 岁对 Judges 仅约 4.6% 的下降（乘子 0.954），但对 FansTotal/FansStar 约 33–34% 的下降（乘子 ≈ 0.66），

且在 FansPerf 上最强（0.637）。因此，年龄主要通过粉丝侧通道（尤其表现诱导与名人底盘分量）发挥作用，而对评委端影响较弱。

5.6 对题问的回答

职业舞者与名人特征均显著影响比赛表现。职业舞者在控制全部名人特征后仍提供显著增量解释力，$\Delta R^2_{\mathrm{pro}}(o)$ 介于 0.052 与 0.099（表 1），且对所有结果都是最重要因素之一（表 2）。名人特征中，年龄总体最强，且对粉丝端的影响显著强于评委端（表 4），行业与州具有中等贡献。这些因素并非通过完全相同机制影响评委与粉丝：类别效应对齐在表现分量上最高、在职业舞者固定盘分量上最低（表 3），说明评委评分最一致地对应于表现驱动变化，而粉丝总份额除表现外还反映固定盘/底盘通道。

5.6.1 带种子重启的两阶段 GA 历史对比

在固定适应度定义不变的前提下，为了对 *dwts_estimator_with_views_modified.py* 中的估计器超参数进行调参，我们对同一 GA 过程运行两次。第一次（Run-1）从随机初始种群出发，用于探索可行区域与更优解区域；第二次（Run-2）采用“种子重启”，其初始种群围绕 Run-1 的最优个体构造，从而更侧重于在局部邻域内进行细化搜索。

表 5–6 给出了两次运行的代表性迭代结果。Run-1 中，最优与平均适应度随迭代提升，同时后验命中概率（`best_hit`）也同步上升。Run-2 中，由于初始即继承 Run-1 的最优解附近信息，起点较高并很快进入平台期，表明主要收益来自于将搜索重心快速迁移至 Run-1 所定位的最优邻域。与此同时，不确定性代理指标（`best_ci`，平均置信区间宽度）与 `best_hit` 的变化并不总是同步，这提示应当以多指标视角对解进行评估，而非仅依赖单一分数。（对应的对比图已以 PNG 形式单独提供。）

局部灵敏度检验（并入本节）。 围绕 GA 得到的最优解 $\theta^\star$，我们采用单因素扰动（one-at-a-time）的方式进行局部鲁棒性检验：对每个参数 $p \in \theta^\star$，令 $p^\pm = p(1 \pm \delta)$

表 5: Run-1（随机初始化）的 GA 历史摘录。

Gen	best_fitness	mean_fitness	best_hit	best_pairprob	best_ci
1	504.953	459.773	0.704	0.767	0.01512
5	542.161	516.606	0.753	0.832	0.01419
15	547.459	523.970	0.771	0.813	0.01358
30	552.248	528.604	0.781	0.812	0.01280

表 6: Run-2（以 Run-1 最优个体为种子重启）的 GA 历史摘录。

Gen	best_fitness	mean_fitness	best_hit	best_pairprob	best_ci
1	552.851	491.709	0.783	0.810	0.01282
6	553.125	543.499	0.784	0.810	0.02022
12	553.125	545.697	0.784	0.810	0.02022
19	553.125	546.522	0.784	0.810	0.02022

（取 $\delta = 0.01$），并重新运行评价流程以计算

$$S_p(y) = \frac{y(\theta^\star \mid p^+) - y(\theta^\star \mid p^-)}{2\delta\, y(\theta^\star)}, \quad y \in \{\texttt{best\_fitness}, \texttt{best\_hit}, \texttt{best\_pairprob}, \texttt{best\_ci}\}.$$

该设计严格对应估计器代码中明确分组的参数结构，包括 (i) 表现投票形状参数（如 `perf_power_excellence`），(ii) 非线性增益/损失形状参数（如 `gain_power_excellence`），(iii) 紧张度映射参数（如 `tight_k`, `tight_threshold`），(iv) 周级反演求解正则项（如 `lam_slack`），以及对数库存更新动力学（如 `star_gain_lr`, `star_decay`）。将 $S_p(\cdot)$ 与上述 GA 历史结果联合呈现，可避免将改进简单归因于单一指标，并使参数扰动与 GA 目标函数中所使用的贝叶斯一致性输出之间形成直接对应关系。

5.6.2 Two-stage GA history (seeded restart)

We run the same GA twice under an unchanged fitness definition. Run-1 starts from a random population for global exploration; Run-2 is a seeded restart initialized around the best individual from Run-1, focusing on local refinement.

Model Sensitivity Analysis. In Run-1, `best_fitness` rises markedly (504.95 →

表 7: Run-1 history excerpt (random init).

Gen	best_fitness	mean_fitness	best_hit	best_pairprob	best_ci
1	504.953	459.773	0.704	0.767	0.01512
16	547.459	528.894	0.771	0.813	0.01358
30	552.248	528.604	0.781	0.812	0.01280

表 8: Run-2 history excerpt (seeded from Run-1 best).

Gen	best_fitness	mean_fitness	best_hit	best_pairprob	best_ci
1	552.851	491.709	0.783	0.810	0.01282
10	553.125	544.789	0.784	0.810	0.02022
19	553.125	546.522	0.784	0.810	0.02022

552.25), accompanied by higher `best_hit` and `best_pairprob` and a smaller `best_ci`, indicating that the model outputs respond substantially to parameter changes at a global scale. In Run-2, `best_fitness` and `best_hit` change only marginally, suggesting local insensitivity of accuracy-type outputs near the optimum; however, `best_ci` can increase while `best_fail_rate`=0, implying that local micro-adjustments mainly affect uncertainty width rather than hit probability. This weighting scheme highlights the model's strong predictive reliability (via `hit` and `pairprob`) while still providing an uncertainty-aware view through `ci`, offering a more informative and well-rounded assessment of performance.