

第二问 (Part 3) 建议：未来赛季推荐使用哪种合成方法？是否保留 Judges' Save？

本 文件对应题目要求：

Based on your analysis, which of the two methods would you recommend using for future seasons and why? Would you suggest including the additional approach of judges choosing from the bottom two couples?

本节基于前两部分的分析结果：

- 跨赛季对比 (Part 1)：同一组投票份额估计下，将 Rank 与 Percent 同时应用到每一季每一周，量化分歧率与“fan-favoring”趋势。
- 争议案例反事实 (Part 2)：对题面指定的争议选手 (Jerry Rice / Billy Ray Cyrus / Bristol Palin / Bobby Bones) 模拟 Rank、Percent 及各自 + Judges' Save 的整季结果。

该小问的目标不是“复现历史节目结果”，而是给出可解释、可操作、与节目定位一致的赛制建议。

1. 记号与规则 (公式原理, 可直接放论文)

设某一周 (固定 season-week) 仍在比赛的选手集合为 $i = 1, \dots, n$ 。

- J_i : 选手 i 当周评委总分 (`total_judge_score`)
- F_i : 选手 i 当周观众投票份额估计 (`fan_vote_normalized` , 满足 $F_i \geq 0$ 且 $\sum_i F_i = 1$)
- 当周真实淘汰人数为 k

1.1 Percent (份额相加, 保留“幅度信息”)

评委百分比：

$$P_i^J = \frac{J_i}{\sum_{j=1}^n J_j}$$

粉丝百分比 (直接使用份额)：

$$P_i^F = F_i$$

合成分 (越小越危险)：

$$C_i^{pct} = P_i^J + P_i^F$$

预测淘汰：取 C^{pct} 最小的 k 人。

1.2 Rank (名次相加, 强调“相对排序”)

定义排名 (1 表示最好) :

- $R_i^J = \text{rank}_{\downarrow}(J_i)$
- $R_i^F = \text{rank}_{\downarrow}(F_i)$

合成排名 (越大越危险) :

$$C_i^{rank} = R_i^J + R_i^F$$

预测淘汰: 取 C^{rank} 最大的 k 人。

1.3 Judges' Save (bottom-2 纠偏机制)

在本项目的反事实模拟中, Judges' Save 的实现口径为:

1. 先按某个基础规则 (Percent 或 Rank) 确定 bottom-2 (最危险 2 人);
2. 在 bottom-2 内淘汰评委分更低者 (J_i 更小);
3. 若当周多淘汰 ($k > 1$), 重复上述过程 k 次。

1.4 算法原理补充: 为什么 Percent 与 Rank 会给出不同结果

这一小节更偏“算法/机制设计”角度, 解释两种合成规则的本质差异: 它们保留的信息不同, 因此在边界周 (分歧周) 会系统性地产生不同淘汰。

(A) Percent = 在同一尺度上做线性加权 (保留幅度)

- Percent 先把评委分变成份额 $P_i^J = J_i / \sum J$, 与粉丝份额 F_i 处于同一尺度, 再相加得到 C_i^{pct} 。
- **关键性质: 保留“幅度信息”。**
 - 例如两位选手粉丝份额差距 0.20 vs 0.02, Percent 会把这种差别直接反映到 C^{pct} 上; 而 Rank 只会把它当作“名次差 1”。
- **尺度不变性 (scale invariance):**
 - 若把所有评委分同时乘上常数 $\alpha > 0$, 则 $J_i / \sum J$ 不变, 因此 Percent 对评委分的单位/计分尺度更稳健。

(B) Rank = 把连续数值压缩为名次 (强调相对排序)

- Rank 先把 J_i 与 F_i 各自转成名次 R_i^J, R_i^F , 再相加。
- **关键性质: 压缩差距、强调排序。**
 - 任意两个相邻名次在 Rank 中“权重相同”, 不管它们的真实差距是很大还是很小。
 - 因而 Rank 对“谁在谁前面”非常敏感, 而对“领先多少”不敏感。
- **不连续性 (discontinuity) 与边界敏感:**

- 当周若存在并列或非常接近的分数/份额，微小波动即可导致名次翻转，从而使 C^{rank} 发生跳变，进而改变 bottom- k 。
- 这就是为什么在很多分歧周里，Rank 更容易把“评委排名靠后”的选手推向淘汰边缘。

(C) 并列 (tie) 与可复现性：为什么需要稳定排序

- 现实数据中经常出现并列分数/并列份额。
- Rank 的实现必须指定并列名次的处理方式（例如 `method='min'`），并在完全相同的合成分/合成名次下使用稳定排序（stable sort）做 tie-break，才能保证输出可复现。
- 这不是“人为操控结果”，而是任何名次型规则在工程实现上都必须面对的细节：一旦进入淘汰边缘（bottom- k ），tie-break 规则就会成为制度的一部分。

(D) Judges' Save 的算法化含义：把“最终淘汰权”局部交给评委

- bottom-2 是一个“候选池”：由基础规则（Percent 或 Rank）筛选出最危险的两对。
- Judges' Save 是在候选池内的二次决策：在 bottom-2 中按评委偏好选择淘汰/拯救。
- 因此它改变了因果链条：
 - 无 Save：观众票与评委分通过合成规则一次性决定淘汰。
 - 有 Save：观众票与评委分首先决定“谁进入 bottom-2”，随后评委在 bottom-2 中决定“谁出局”。
- 这解释了为什么 Save 往往会降低争议型结果：对“评委弱但粉丝强”的选手而言，即使粉丝票能把他推离最危险位置，一旦仍进入 bottom-2，就会在第二阶段面临更高的淘汰概率。

2. 用于“推荐”的证据：我们看什么、为什么看它

2.1 “更偏向观众投票”的量化：fan-favor proxy

在每个淘汰周内部，把粉丝份额和评委分数都转为分位数（ascending，越低越危险）：

- $Q_i^F = \text{pct-rank}_{\uparrow}(F_i)$
- $Q_i^J = \text{pct-rank}_{\uparrow}(J_i)$

对某规则预测的淘汰集合 E （大小为 k ），计算：

$$\Delta = \mathbb{E}_{i \in E} [Q_i^F - Q_i^J]$$

解释：

- 若 Δ 更负：淘汰更像由“粉丝票低”驱动（更 fan-driven）。
- 若 Δ 接近 0 或为正：淘汰更像由“评委分低”驱动（更 judge-driven）。

这类指标之所以重要，是因为“推荐赛制”本质上是价值选择：节目到底是更强调观众参与决定性，还是更强调评委专业性与争议抑制。

2.2 “争议案例”检验：制度在极端情况下会怎样

题面给的四个例子本质上是在问：当出现 judge-fan divergence 时，制度会不会导致“高争议结果”（例如低评委认可者走到很后甚至夺冠）。

因此我们把这四个案例作为“压力测试”：看 Rank / Percent / +Save 是否会显著改变淘汰周与最终名次。

3. 图表（本问专用）与读图说明

本文件配套输出两张图，均位于本目录：`recommendation_outputs/figures/`。

图1：fig_fan_favor_proxy_percent_vs_rank.png

图名（中文建议）：Percent 与 Rank 在全部淘汰周的 fan-favor 代理指标分布对比 ($\Delta = \mathbb{E}[Q^F - Q^J]$)。

Figure caption (English): Distribution of the fan-favor proxy across all elimination weeks under Percent vs Rank, where $\Delta = \mathbb{E}[Q^F - Q^J]$ is computed among predicted eliminations.

它说明什么（你要写进论文的要点）：

- **如何读这张图：**

- 每个点代表一个淘汰周（season-week），在该周内按某规则得到预测淘汰集合 E ，再计算 $\Delta = \mathbb{E}_{i \in E} [Q_i^F - Q_i^J]$ 。
- Q^F, Q^J 都是“从小到大”的分位（越小越危险），因此 $\Delta < 0$ 表示：被淘汰者在粉丝端更危险、而在评委端相对不那么危险，即淘汰更像由“粉丝票低”驱动。

- **核心观察 1：Percent 更偏向 fan-driven。**

- Percent 的 Δ 分布整体更偏负，意味着它更倾向让“粉丝份额更低的人”出局。
- 这在制度解释上对应：Percent 直接使用份额幅度 F_i ，因此当某选手粉丝票显著偏低时会更稳定地落入淘汰边缘。

- **核心观察 2：Rank 更偏向 judge-driven。**

- Rank 的 Δ 更靠近 0（甚至部分周为正），意味着它更容易淘汰“粉丝并不低但评委排名差”的选手。
- 这在制度解释上对应：Rank 把连续差距压缩为名次，相当于强化了“相对排序”而弱化了“领先幅度”，因此评委端的相对落后更容易直接传导到淘汰风险。

- **这张图对“推荐”的直接含义：**

- 如果节目目标是“让观众投票更具决定性”，那么希望 Δ 更负（淘汰更由粉丝票低驱动），因此 Percent 更符合该目标。
- 如果节目目标是“让评委排序更强约束结果、减少争议型走得太远”，那么 Rank 的分布特征更贴近该目标。

- **常见误读澄清（建议写进 discussion）：**

- 图1不是在说某规则“更正确/更贴合历史淘汰”，而是在比较两种规则的制度倾向：它们在淘汰边缘更倾向“惩罚粉丝低”还是“惩罚评委低”。

论文可直接引用的句式（中文）：

Figure X 显示在所有淘汰周中，Percent 合成规则下的 $\Delta = \mathbb{E}[Q^F - Q^J]$ 分布整体更偏负，表明其预测淘汰更倾向于由“粉丝支持不足”驱动；相比之下，Rank 的 Δ 更接近 0，意味着其更可能淘汰粉丝支持并不低但评委排名靠后的选手，从而呈现更强的评委主导倾向。

目的：为“推荐哪种合成方法”提供跨赛季、跨周的总体证据，而不是依赖单一案例。

图2：fig_case_counterfactual_final_place.png

图名（中文建议）：争议案例在四种赛制（Actual / Percent / Rank / Percent+Save / Rank+Save）下的反事实最终名次对比（名次 1 最好）。

Figure caption (English)：Counterfactual season outcomes for controversy cases under four mechanisms (Actual, Percent, Rank, Percent+Save, Rank+Save). Lower final place indicates better performance.

它说明什么（你要写进论文的要点）：

- **如何读这张图：**

- 横轴是四个争议案例（S2 Jerry、S4 Billy、S11 Bristol、S27 Bobby）。
- 每个案例有 5 根柱：Actual（真实名次）与 4 个反事实情景（Percent、Rank、Percent+Save、Rank+Save）。
- 纵轴是最终名次（1 最好），图中做了反转（更靠上代表名次更差），因此“柱子向上”表示结果变差。

- **核心观察 1：规则选择会改变赛季级结果，且对争议型选手影响更大。**

- 对典型争议选手（尤其 Bobby Bones、Bristol Palin），Rank 会显著压低最终名次；Percent 相对更“允许粉丝强支撑”抵消评委劣势。
- 这说明两种规则不仅会影响“某一周淘汰谁”，还会沿着淘汰路径累积，最终改变赛季结果（甚至改变冠军归属）。

- **核心观察 2：Judges' Save 的作用方向一致——把决定权在 bottom-2 处向评委倾斜。**

- Save 并不重新定义整周的合成方法，它只在 bottom-2 触发；因此它更像一个“边缘纠偏/安全阀”。
- 一旦争议型选手进入 bottom-2（通常意味着评委端相对更弱），Save 会提高其被淘汰的概率，从而把争议从“冠军/四强层面”压回到更早周次。

- **这张图对“推荐”的直接含义：**

- 如果节目希望保留观众票的总体决定性，同时避免极端争议结果，那么“Percent + Save”提供了折中：
 - 平时用 Percent 保留份额幅度与观众参与；
 - 仅在 bottom-2 由 Save 进行局部纠偏，防止少数极端路径导致高争议结局。

论文可直接引用的句式（中文）：

Figure Y 将四个争议案例在不同赛制下的最终名次并列展示。结果表明，Rank 规则会显著压低典型争议型选手（评委端弱、粉丝端强）的赛季成绩，而 Percent 更允许粉丝优势抵消评委劣势；进一步引入 Judges' Save 后，两种规则都会在 bottom-2 阶段增强评委影响，从而系统性降低争议型选手走到最后的概率。

目的：展示“制度选择会改变赛季级结果”的量级，并回答题面关于 Judges' Save 的制度影响。

4. 推荐（本小问的最终答案，可直接放结论）

4.1 推荐的合成方法：Percent

建议未来赛季优先采用 Percent（份额相加）作为主合成规则。理由：

1. 尺度一致且保留幅度信息： $J / \sum J$ 与 F 都是份额，直接相加具有清晰解释；并能区分“领先很多”和“领先一点”的差异。
2. 更符合观众投票决定性的节目定位：跨赛季统计显示 Percent 的淘汰更偏向“粉丝支持较低者出局”（ Δ 更负）。
3. 边界更平滑：Rank 把连续数值压缩为名次，遇到并列/接近时更容易产生跳变；Percent 相对更连续。

4.2 是否建议加入 Judges' Save：建议保留（作为 bottom-2 的安全阀）

建议保留 Judges' Save，但明确其定位是“争议抑制的安全阀”，而不是替代观众投票。

- Save 只在 bottom-2 介入，是一种局部纠偏：它不会每周都把决定权交给评委，但会在最危险的边缘周次提高评委约束。
- 争议案例反事实显示，Save 会系统性降低“评委弱但粉丝强”选手走到最后的概率，从而减少高争议结果。

4.3 一句话结论（建议放摘要/讨论段）

若节目目标强调观众参与度与“票数幅度”对结果的影响，推荐采用 Percent 作为主合成规则；同时保留 Judges' Save 作为淘汰边缘的纠偏机制，以降低极端争议结果（例如低评委认可者夺冠）发生概率。

5. 复现与文件索引

- 生成图：运行 `recommendation_outputs/build_recommendation_figures.py`
- 图文件：
 - `recommendation_outputs/figures/fig_fan_favor_proxy_percent_vs_rank.png`
 - `recommendation_outputs/figures/fig_case_counterfactual_final_place.png`

上游数据来自：

- `method_compare_outputs2.1/weekly_method_comparison.csv`
- `controversy_outputs/case_summary.csv`