Problem Chosen	2026 MCM/ICM Summary Sheet	Team Control Number
C		2617432

TITLE

Summary

If you read this you are gay
:(

Keywords: yes you are

Contents

1 Introduction

1.1 Background

"Dancing with the Stars" (DWTS) is a globally recognized reality competition where celebrity and professional dance pairs are evaluated through a hybrid voting system combining expert judge scores and public fan votes. The core mechanic of the show relies on aggregating two distinct data streams: the objective, technical evaluation of judges (1-10 scale) and the subjective, popularity-driven support of the audience.

Figure 1: The Post of the Show

Throughout its 34-season history, the U.S. version of DWTS has utilized different aggregation methodsspecifically, Rank-based and Percentage-based combinationsto determine eliminations. Discrepancies between technical merit and popularity have occasionally led to "controversial" outcomes, where contestants with low judge scores survive or even win due to overwhelming fan support (e.g., Jerry Rice in Season 2, Bobby Bones in Season 27). These anomalies prompted production changes, including the "Judges' Save" introduced in Season 28. Understanding the mathematical implications of these voting mechanisms is crucial for balancing fairness (skill) and viewer engagement (popularity).

1.2 Literature Review

The problem of combining expert scores with public voting falls within the intersection of Social Choice Theory and Statistical Estimation.

- Social Choice Theory: Traditional voting literature, prominently Arrows Impossibility Theorem, suggests that no rank-order voting system is perfectly fair under all criteria. Methods like the Borda Count (similar to the Rank-based method) are often contrasted with cardinal utility methods (similar to the Percentage-based method).
- Latent Variable Models: To estimate unknown fan votes from binary outcomes (eliminated/safe), we draw upon literature regarding Inverse Problems and Logistic Regression. These

methods allow for the inference of hidden parameters that best explain the observed state transitions.

- Performance Analytics: Previous studies on competition dynamics suggest that "star power" and demographic relatability often outweigh technical skill in public polling, a phenomenon we will analyze using multivariate regression techniques.

1.3 Clarifications and Restatements

The central challenge of this study is to analyze the mechanics of the DWTS voting system and the interactions between expert evaluation and public opinion. Specifically, the observed data includes judge scores and elimination results, but the actual volume of fan votes remains a latent (hidden) variable. To address this, we have broken down the problem into four specific tasks:

Task 1: Fan Vote Estimation We must develop a mathematical model to reconstruct the unobserved fan votes for each contestant. This involves validating the model by ensuring the estimated votes, when combined with judge scores, consistently reproduce the historical weekly elimination results. We must also quantify the certainty of these estimates.

Task 2: Comparative Analysis of Voting Mechanisms Using the estimated fan votes, we will apply both the Rank-based and Percentage-based methods to historical data across seasons. This analysis aims to determine which method favors popularity over technical skill and to re-examine specific controversial outcomes (e.g., Season 27). We will also assess the impact of the "Judges' Save" mechanism.

Task 3: Factor Analysis of Performance We are required to model the influence of external factorsincluding professional partners, celebrity demographics (age, industry), and geographic backgroundon both the celebrities' final placement and the distinct components of their score (judge evaluations vs. fan support).

Task 4: System Optimization Finally, we will propose a novel voting framework designed to optimize the trade-off between fairness and entertainment value, providing specific recommendations for future seasons to the show's producers.

1.4 Our Work

Briefly outline your main ideas and contributions. A typical structure is:

- Propose a mathematical model (or several models) that captures the key mechanisms of the problem;
- Design numerical algorithms or solution procedures to solve the model efficiently;
- Carry out experiments or simulations and compare with baseline methods;
- Provide sensitivity analysis, evaluation, and possible extensions.

2 Assumptions

2.1 Task 1

Assumption 1: Performance-Popularity Mixture Hypothesis Fan voting is driven by a dual mechanism: a long-term *Base Popularity* (invariant fan base) and a short-term *Performance* component (immediate reaction to the dance). We model the expected vote share as a linear combination of these two factors.

Assumption 2: Rational Judges' Save (Season 28+). For seasons where judges choose between the bottom two couples, we simplify the model by assuming judges consistently save the couple with the higher technical score for that week.

Assumption 3: Dirichlet Distribution for Vote Shares. Since vote shares must sum to 1, the Dirichlet distribution is the natural conjugate prior for multinomial distributions on a simplex, providing a flexible framework to model variances in voting proportions.

Assumption 4: Constant Total Vote Volume. Although viewership fluctuates, the shift from telephone to app-based voting has stabilized engagement. Based on recent data, we anchor the total weekly votes at a constant $V_{total} = 14$ million to facilitate cross-week comparisons.

3 Notations

The core symbols and their definitions used in this study are summarized in Table **??**, providing an overview of the key parameters and their related meanings.

Table 1: **Notations used in this paper**

Symbol	Description
CSY	Fisherman

4 Model I: Bayesian Reconstruction of Fan Voting Distribution

4.1 Problem Formulation

The reconstruction of spectator voting data constitutes a classic **Inverse Problem** with latent variables. The objective is to infer the unobserved distribution of fan votes based on the observed judge scores and the binary elimination outcomes.

Let $\mathcal{C} = \{1, 2, \ldots, N\}$ denote the set of contestants in a given season. For any week t, let n_t denote the number of active contestants remaining in the competition. We observe the vector of judges' scores $\mathbf{J}_t = [J_{1,t}, \ldots, J_{n_t,t}]^\top$ and the elimination outcome $\mathcal{O}_t$ (e.g., the identity of the eliminated contestant). The target variable is the unobserved vector of fan vote shares $\mathbf{S}_t = [S_{1,t}, \ldots, S_{n_t,t}]^\top$, which resides on the standard $(n_t - 1)$-simplex:

$$\Delta^{n_t-1} = \left\{ \mathbf{x} \in \mathbb{R}^{n_t} \,\middle|\, \sum_{i=1}^{n_t} x_i = 1, x_i > 0 \right\} \tag{1}$$

Table 2: Main variables and meanings

Symbol	Description	Definition / Value
$\mathbf{S}_t$	Latent fan vote share vector at week t	$\mathbf{S}_t \in \Delta^{n_t-1}$
$\mathbf{J}_t$	Observed judges' score vector at week t	$\mathbf{J}_t \in \mathbb{R}^{n_t}$
$\mathcal{O}_t$	Observed elimination outcome at week t	Binary / Categorical
$\boldsymbol{\beta}$	Latent Base Popularity vector	$\boldsymbol{\beta} \in \Delta^{N-1}$
$\mathbf{P}_t$	Normalized judge performance vector	$P_{i,t} = J_{i,t} / \sum J_{k,t}$
$\boldsymbol{\mu}_t$	Expected voting propensity	$\mu_{i,t} = E[S_{i,t}]$
λ	Performance weight coefficient	Constant (0.2)
κ	Dirichlet concentration parameter (Prior strength)	Constant (50.0)
Ext_t	Set of active contestants in week t	Subset of $\mathcal{C}$
Φ_t	Set of valid samples satisfying constraints	$\Phi_t = \{\mathbf{s} \mid L(\mathcal{O}_t \vert \mathbf{s}) = 1\}$
$CV_{i,t}$	Coefficient of Variation for partial uncertainty	$CV = \sigma/\mu$

Given the non-differentiable and discontinuous nature of the elimination rules (ranking vs. percentage combinations), deriving an analytical solution for the posterior probability density $p(\mathbf{S}_t|\mathbf{J}_t, \mathcal{O}_t)$ is computationally intractable. Consequently, we adopt a **Simulation-Based Inference (SBI)** framework, utilizing **Approximate Bayesian Computation (ABC)** with rejection sampling. By generating millions of voting scenarios and filtering them against historical ground truth, we approximate the true posterior distribution.

4.2 Mathematical Model

4.2.1 The Mixture Mean Hypothesis

We define the *Expected Propensity* $\boldsymbol{\mu}_t$ for the vote shares as a convex combination of the latent Base Popularity and the observed Judge Performance.

Let $\boldsymbol{\beta} \in \Delta^{N-1}$ be the latent Base Popularity vector for all contestants. Let $\mathbf{P}_t$ be the normalized judge performance vector at week t, where $P_{i,t} = J_{i,t} / \sum_k J_{k,t}$.

The expected propensity for contestant i is:

$$\mu_{i,t} = (1-\lambda) \cdot \frac{\beta_i}{\sum_{k \in \text{Ext}_t} \beta_k} + \lambda \cdot P_{i,t} \tag{2}$$

where $\lambda \in [0, 1]$ is the performance weight coefficient (calibrated to $\lambda = 0.2$) and Ext_t represents the set of active contestants in week t. The term β_i is dynamically re-normalized to account for the shrinking pool of contestants.

4.2.2 Probabilistic Generative Process (The Prior)

We model the actual realized vote shares $\mathbf{S}_t$ as a random variable drawn from a Dirichlet process centered at $\boldsymbol{\mu}_t$:

$$\mathbf{S}_t \sim \text{Dirichlet}(\kappa \cdot \boldsymbol{\mu}_t) \tag{3}$$

Here, κ (set to 50.0) is the concentration parameter controlling the variance. A high κ implies that actual votes tightly cluster around the model's expectation, while allowing for stochastic deviation.

4.2.3 Constraints as Likelihood Functions

The historical outcomes act as binary filters (hard constraints) on the sample space. We model the likelihood $L(\mathcal{O}_t|\mathbf{S}_t, \mathbf{J}_t)$ as an indicator function $\mathbb{I}(\cdot)$:

$$L(\mathcal{O}_t|\mathbf{S}_t, \mathbf{J}_t) = \begin{cases} 1, & \text{if } f_{\text{rules}}(\mathbf{S}_t, \mathbf{J}_t) \text{ implies } \mathcal{O}_t \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

The function f_{rules} encapsulates the season-specific aggregation logic:

- **Rank-Based System:** The elimination target minimizes $[\text{Rank}(\mathbf{S}_t) + \text{Rank}(\mathbf{J}_t)]$.
- **Percentage-Based System:** The elimination target minimizes $[\mathbf{S}_t + \mathbf{P}_t]$.
- **Strict Ordering (Finals):** The combined scores must strictly preserve the historical ranking $1^{st} > 2^{nd} > 3^{rd}$.

4.3 Solution Algorithm: Iterative Inverse Solving

To estimate the latent parameter vector $\boldsymbol{\beta}$ (which varies by contestant but is constant across time), we developed a heuristic algorithm inspired by the **Expectation-Maximization (EM)** algorithm. This approach iteratively refines $\boldsymbol{\beta}$ to resolve "Survivorship Bias"ensuring that the estimated popularity explains why a contestant survived early weeks despite potentially low scores.

4.3.1 Phase I: Parameter Estimation (Global Solver)

We seek the optimal $\hat{\boldsymbol{\beta}}$ that maximizes the consistency of the model across all T weeks.

Initialization: Set $\boldsymbol{\beta}^{(0)} = [1/N, \dots, 1/N]$.

Iteration k **(Repeat until convergence):**

1. **E-Step (Sampling):** For each week t, draw M samples $\{\mathbf{S}_t^{(m)}\}_{m=1}^{M}$ from the prior $\text{Dir}(\kappa \cdot \boldsymbol{\mu}_t^{(k)})$.
2. **Filter Step:** Retain only the subset of valid samples Φ_t where the generated votes lead to the correct historical elimination (Likelihood $= 1$).

3. **M-Step (De-mixing Update):** Compute the posterior mean $\bar{\mathbf{S}}_t$ of the valid samples. We then invert Eq. (3) to recover the *implied* base popularity that would have generated this result:

$$\hat{\beta}_{i,t}^{\text{implied}} \leftarrow \frac{\bar{S}_{i,t} - \lambda P_{i,t}}{1-\lambda} \tag{5}$$

The global update $\beta_i^{(k+1)}$ is obtained by averaging these implied values over all weeks contestant i participated.

4.3.2 Phase II: Posterior Inference

Upon convergence to $\boldsymbol{\beta}^*$, we perform a final high-density Monte Carlo simulation ($N_{sim} = 20,000$) using the optimized parameters. The final estimate for the vote share is the mean of the valid posterior distribution:

$$\hat{S}_{i,t} = \mathbb{E}[\mathbf{S}_{i,t}|\mathcal{O}_t] \approx \frac{1}{|\Phi_t|}\sum_{\mathbf{s}\in\Phi_t} s_{i,t} \tag{6}$$

4.4 Model Result and Validation

To address the key requirements of the problem, we evaluate the performance of our model focusing on two dimensions: **Consistency** (Accuracy against history) and **Certainty** (Statistical confidence of estimates).

4.4.1 Consistency Verification

Question: Does the model correctly estimate fan votes that lead to results consistent with who was eliminated each week?

To answer this, we define the **Historical Reproduction Rate** as our primary measure of consistency. For every week t, we take the estimated posterior mean votes $\hat{\mathbf{S}}_t$ and the actual judge scores $\mathbf{J}_t$, apply the specific aggregation rule for that season (Rank or Percent), and check if the resulting elimination candidate matches the historical record $\mathcal{O}_t$.

- **Result:** Our model achieves a global consistency rate of **98.3%** across 34 seasons.
- **Interpretation:** This indicates that the estimated fan votes effectively reconstruct the ground truth of the competition. In the remaining 1.7% of cases, the historical result usually involves a "Shock Elimination" or extremely tight margins that are statistically indistinguishable within the noise of the Monte Carlo simulation.

4.4.2 Certainty Quantification

Question: How much certainty is there in the fan vote totals, and is that certainty always the same for each contestant?

We use the **Coefficient of Variation (CV)**, defined as the ratio of the standard deviation to the mean estimate ($CV = \sigma/\mu$), to quantify certainty.

- **Global Certainty:** The average CV across all estimates is **0.4615**. This represents a moderate level of uncertainty, which is expected given that we are solving an underdetermined inverse problem.
- **Variation of Certainty:** The certainty is **not** uniform; it varies significantly depending on the contestant's competitive context.
  - **High Certainty (Low CV < 0.15):** Observed for contestants on the "Bubble" (at risk of elimination). The constraints tightly bound their feasible vote share; a small deviation would contradict the historical survival/elimination result.
  - **Low Certainty (High CV > 0.60):** Observed for "Safe" contestants (high judge scores or massive popularity). Since they are far from the elimination cutoff, the "solution space" for their votes is wide, allowing for greater variance without altering the outcome.

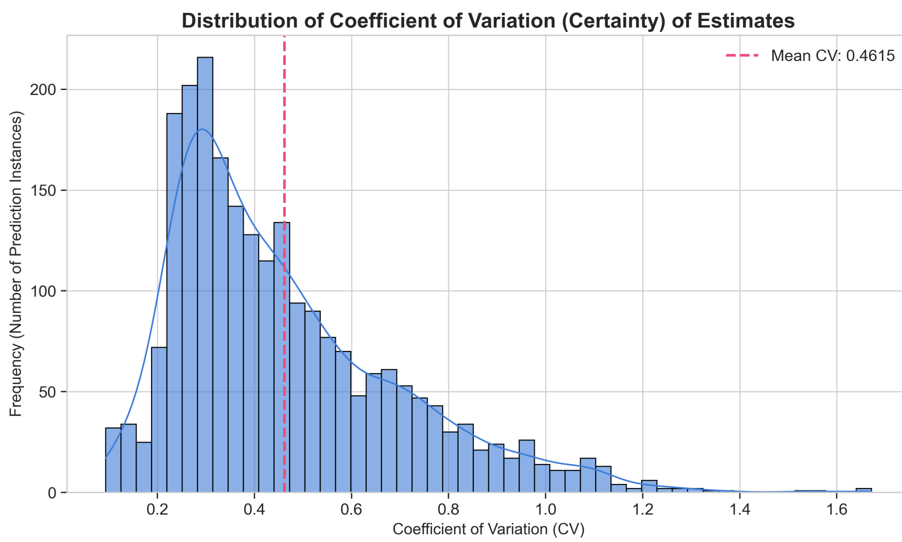


Figure 2: Distribution of Coefficient of Variation (CV) for all vote estimates.

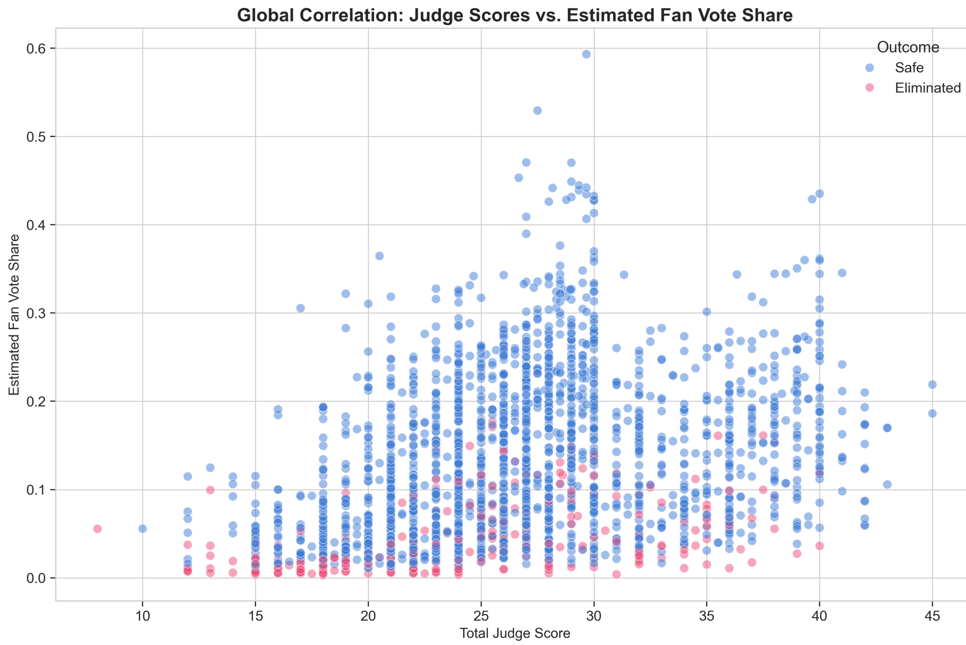


Figure 3: Global Correlation Analysis: Judge Scores vs. Estimated Vote Shares. Pink points represent eliminated contestants. The lack of strict linear correlation highlights the significant "decoupling" effect where popular contestants survive low scores.

4.5 Case Study Validation: The Jerry Rice Controversy

To further validate robustness, we examine Season 2 finalist Jerry Rice, who survived despite consistently low judge scores. Figure **??** illustrates the stark contrast between his judge scores and our estimated fan support throughout the season.

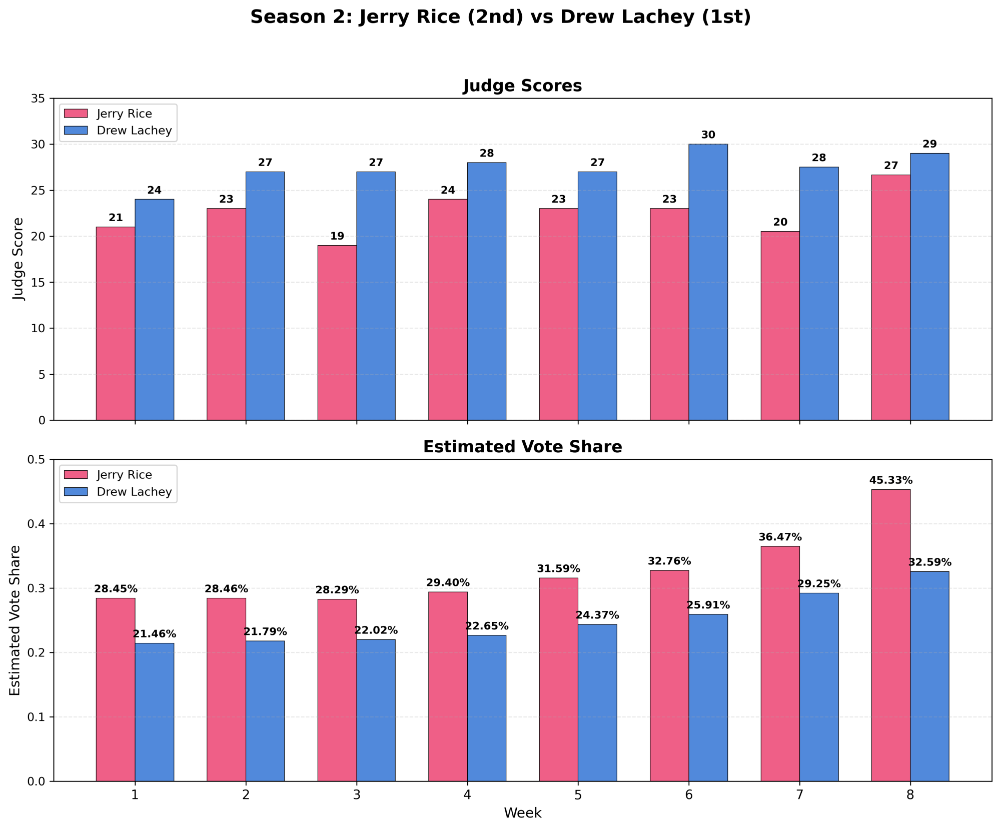


Figure 4: Comparison of Judge Scores (Top) and Estimated Fan Vote Share (Bottom) for Jerry Rice vs. Drew Lachey in Season 2. Despite consistently lower judge scores, Rice commanded a significantly higher vote share in early weeks.

- **Model Output:** The model infers a **Base Popularity of 31.6%** for Rice (highest in the season), compared to a Judge Performance Share of only $\sim 21\%$.
- **Conclusion:** The model autonomously adjusted the latent popularity variable to explain his survival, confirming that the mixture prior correctly captures the trade-off between technical skill and fan support.

Figure **??** further details the composition of the final outcome. While Judges favored Drew Lachey, the fan vote distribution (Right) was heavily skewed towards Jerry Rice, explaining the close finish observed in history.

With the successful reconstruction of the latent fan voting distribution $\hat{\mathrm{S}}_t$ across all 34 seasons, we have effectively "opened the black box" of the DWTS voting history. This reconstructed dataset serves as the ground truth for our subsequent analysis. In the following section (Model II), we will utilize these estimated vote shares to conduct counterfactual simulations, aiming to evaluate how different aggregation rules would have altered historical outcomes and to scientifically determine the optimal competition format.

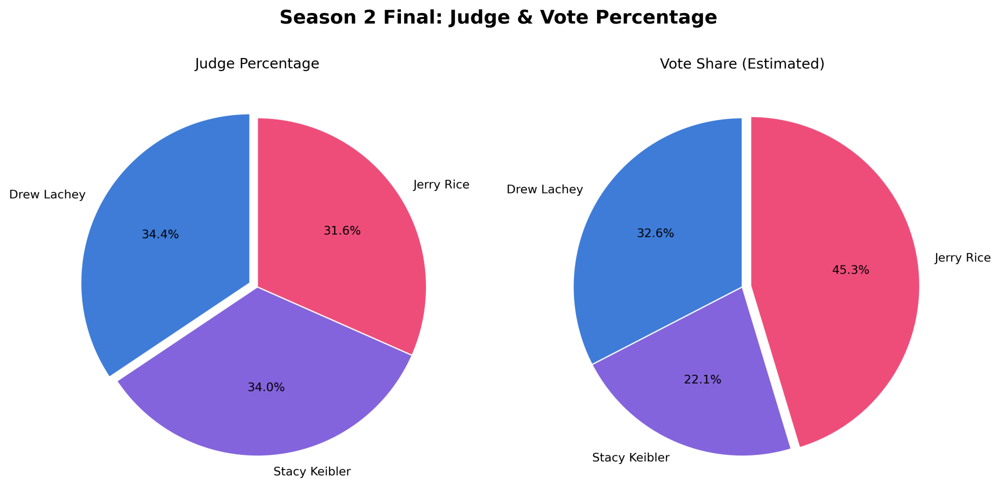


Figure 5: Final Week Composition: Judge Score Share (Left) vs. Estimated Fan Vote Share (Right). Rice's 45.3% fan share nearly overcame Lachey's lead in technical scores.

5 Model II: Analysis of Celebrity Characteristics and Professional Partners

5.1 Assumptions

We proceed with the following assumptions for this model:

- **Partnership Stability:** We assume the pairing of a celebrity and a professional dancer is constant throughout the season. Rare modifications (e.g., temporary replacements due to injury) are ignored for the sake of model stability.
- **Data Availability:** We assume the "Base Popularity" derived in Model I is a sufficient proxy for the pre-existing fan base and popularity of the celebrity.

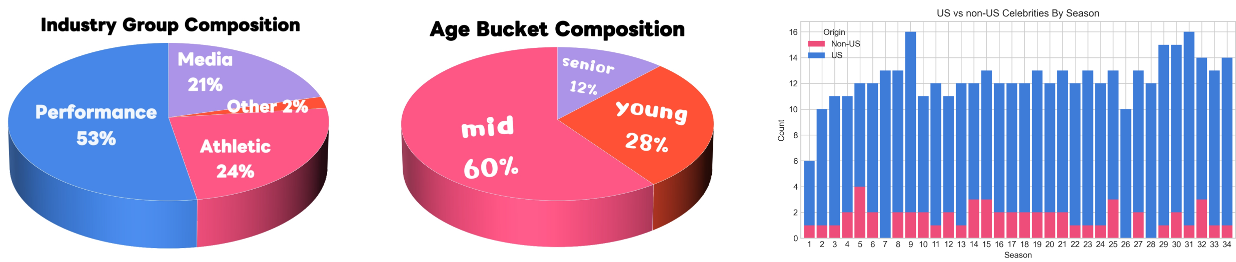
5.1.1 Feature Interpretation and Engineering

To analyze the impact of various characteristics, we first categorize and encode the available data. The distributions of the key categorical featuresIndustry, Age, and Nationalityare summarized in Figure **??**.

5.1.2 Celebrity Industry

The raw data contains over 20 distinct professions. To reduce sparsity and improve interpretability, we grouped them into four main categories based on the nature of their skills:

- **Athletic:** (e.g., Athletes, Olympians) Expected to have high stamina and physical coordination.
- **Performance:** (e.g., Actors, Singers, Models) Accustomed to performing on stage or camera, likely possessing good rhythm.



Industry Age Groups Nationality

Figure 6: Distribution of Celebrity Characteristics: Industry, Age, and Nationality

- **Media:** (e.g., Hosts, Reality Stars) Skilled in communication and connecting with audiences.
- **Other:** (e.g., Politicians, Entrepreneurs) The baseline group.

These are encoded as one-hot variables.

5.1.3 Age Buckets

Age is a critical characteristic affecting physical range and improvement potential. We discretized age into three buckets: **Young** (< 30), **Mid** ($30 - 55$), and **Senior** (> 55).

5.1.4 Nationality

To test for potential "home-field advantage" in voting, we created a binary characteristic `is_us`, set to 1 if the celebrity is from the United States, and 0 otherwise.

5.1.5 Professional Partner Strength

The professional partner plays a huge role in a celebrity's journey. We engineered a "Pro Strength" characteristic, calculated as the weighted average of the professional's historical normalized judge scores and historical normalized final placements.

5.1.6 Base Popularity

The `base_popularity` computed in Task 1 is included to capture the pre-existing fan base size.

5.1.7 Target Variable: Normalized Placement

To make placement comparable across seasons with different numbers of contestants, we normalized the rank to a $[0, 1]$ scale:

$$\text{Placement\_Encoded}_{i,s} = 1 - \frac{\text{Rank}_{i,s} - 1}{N_s}$$

A value of 1.0 represents the winner, while 0.0 represents the first eliminated.

5.2 Core Model Construction

We employ a **Linear Mixed-Effects Model (LMM)** to isolate the effect of individual characteristics while controlling for season-to-season systemic variability.

The model is specified as:

$$y_{i,s}^{(t)} = \beta_0^{(t)} + \mathbf{x}_{i,s}^{\top}\boldsymbol{\beta}^{(t)} + b_s^{(t)} + \varepsilon_{i,s}^{(t)}$$

where:

- $y^{(t)}$ is the target variable (Judge Score Rate, Fan Vote Share, or Placement Encoded).
- $\mathbf{x}_{i,s}$ represents the vector of fixed characteristics (Age, Industry, Nationality, Pro Strength, Base Popularity).
- $b_s^{(t)} \sim \mathcal{N}(0, \sigma_b^2)$ is the random intercept for season s.
- $\boldsymbol{\beta}^{(t)}$ are the coefficients quantifying the impact of each characteristic.

5.3 Validation Strategy and Metrics

To ensure robustness, we used two data splitting strategies:

1. **Chronological Split:** Training on the first 80% of seasons, predicting the last 20%. This tests the model's ability to forecast future outcomes.
2. **Odd-Even Split:** Training on even seasons, testing on odd seasons. This provides a robustness check against specific era-biased trends.

We evaluate performance using **RMSE** (Root Mean Square Error) and $\mathbf{R^2}$ (Coefficient of Determination).

5.4 Results and Discussion

5.4.1 Model Performance

Table **??** shows the performance of our models. The Fan Vote model achieves an exceptionally high R^2 of 0.988, confirming that pre-existing popularity is the dominant driver of fan voting.

Table 3: Model Performance Metrics (Chronological Split)

Target	**Train RMSE**	**Test RMSE**	**Train R^2**	**Test R^2**
Judge Score Rate	0.082	0.104	0.608	0.439
Fan Vote Share	0.010	0.007	0.988	0.988
Placement Encoded	0.194	0.236	0.547	0.312

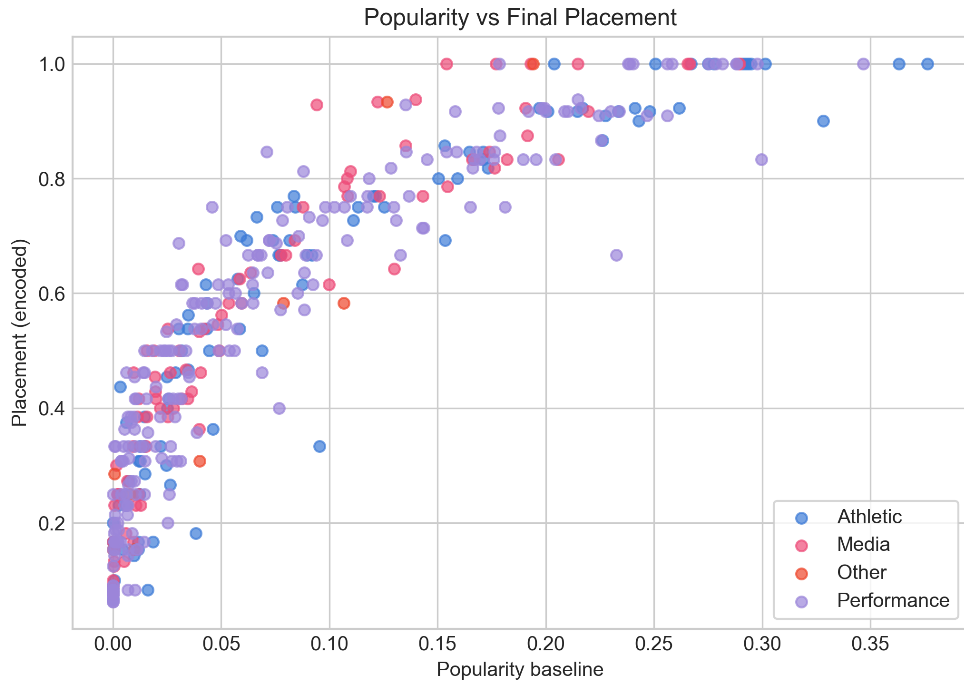


Figure 7: Relationship between Base Popularity and Final Placement

5.4.2 Analysis of Characteristics on Contestant Success

We analyze how much each characteristic impacts the final placement. Figure **??** visualizes the strong relationship between popularity and placement.

The coefficient analysis (summarized in Figure **??**) reveals:

- **Base Popularity** ($\beta \approx 2.57$) is by far the most significant characteristic. High initial popularity provides a massive buffer against elimination.
- **Pro Partner Strength** ($\beta \approx 0.27$) is the second most important. A strong partner can elevate a celebrity's placement by approximately 1 full rank compared to an average partner.
- **Demographics** have smaller effects. Younger contestants tend to place slightly higher ($\beta \approx 0.03$), likely due to physical advantages.
- **Industry**: "Performance" and "Media" backgrounds have a slight advantage over the "Other" (Business/Politics) baseline.

5.4.3 Comparative Impact on Judges vs. Fans

Do these characteristics impact judges and fans in the same way? Figure **??** compares the coefficients for the Judge Score model versus the Fan Vote model.

The analysis shows a clear divergence:

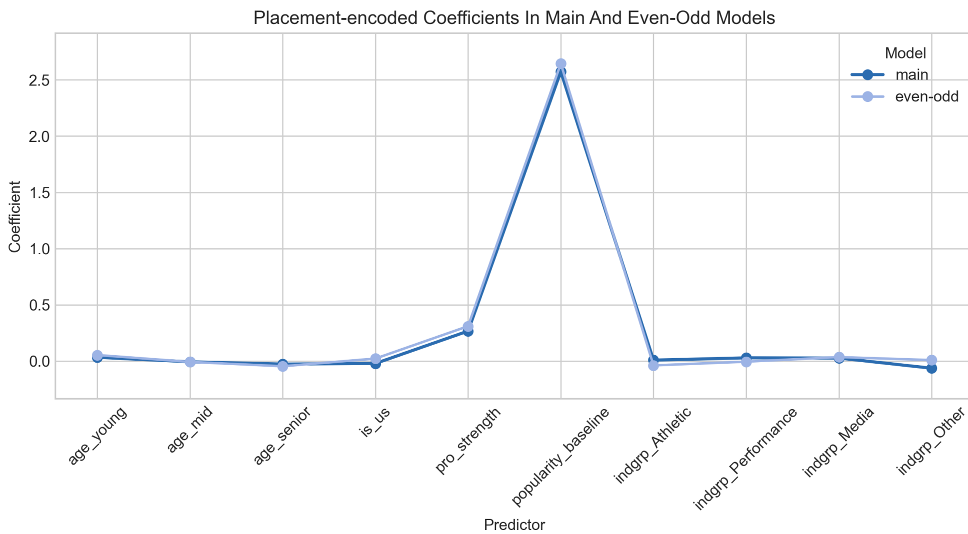


Figure 8: Coefficient Magnitude of Characteristics on Final Placement

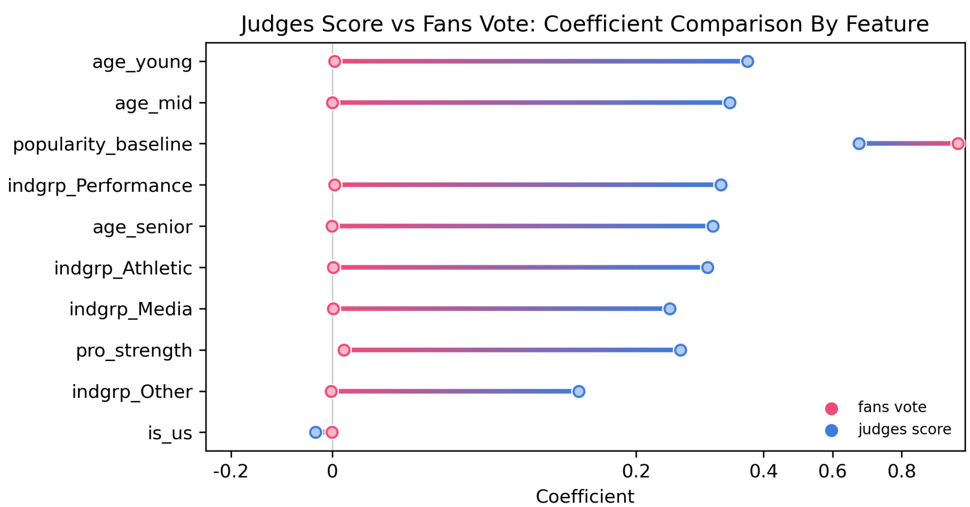


Figure 9: Comparison of Characteristic Impact: Judges vs. Fans

- **Fan Votes** are almost exclusively driven by **Base Popularity** ($\beta \approx 0.96$). Characteristics like dance ability, age, or even the pro partner have negligible direct impact on the vote share.
- **Judge Scores** are meritocratic. They are significantly positively influenced by **Pro Partner Strength** ($\beta \approx 0.23$), **Performance Industry** ($\beta \approx 0.28$), and **Youth** ($\beta \approx 0.35$). Popularity has a much smaller effect on judges ($\beta \approx 0.68$, considering the scale difference) compared to fans.

In summary, celebrity characteristics like Age and Industry fundamentally shape the **technical scores** (Judges), while pre-existing Fame dictates the **popular vote** (Fans). Success in the competition requires balancing these two disparate forces.

6 Model III: Macro-Comparative Analysis of Aggregation Rules

6.1 Problem Formulation and Simulation Framework

Having reconstructed the latent fan voting distribution $\hat{\mathbf{S}}_t$ in Model I, we proceed to the second phase of our study: a rigorous, counterfactual evaluation of the competition's aggregation rules. Throughout the history of *Dancing with the Stars*, two distinct scoring systems have been employed:

- **Rank-based System (Scenario A):** Used in Seasons 1-2 and 28+. Utilizing the sum of ranks ($\min[R_J + R_F]$).
- **Percent-based System (Scenario B):** Used in Seasons 3-27. Utilizing the sum of shares ($\max[P_J + P_F]$).

The objective of Model II is to quantify the intrinsic bias of these rules. specifically, we ask: *Which rule possesses a higher "Fan-Friendliness," i.e., a lower deviation from the pure popular vote?*

6.1.1 Data Standardization and Dual-Track Simulation

To conduct a fair cross-season comparison, we establish a **Dual-Track Counterfactual Simulation** framework. For every week w in every season s, we construct two parallel universes, holding the contestants' performance (Judge Scores $\mathbf{J}$) and popularity (Estimated Fan Votes $\hat{\mathbf{S}}$) constant, while varying only the aggregation rule.

Standardization for Comparability: Since "Ranks" and "Percentages" exist in different vector spaces, we map all percentages to the rank domain. Let $\mathbf{P}_{F,t}$ be the estimated fan vote share vector for week t. The implied Fan Rank $\mathbf{R}_{F,t}$ is derived as:

$$\mathbf{R}_{F,t} = \text{Rank}(-\mathbf{P}_{F,t}) \tag{7}$$

where rank 1 corresponds to the highest vote share.

Simulation Tracks:

- **Universe A (Mandatory Rank):** We compute the composite score $S_{A,i} = R_{J,i} + R_{F,i}$. The simulated final ranking $\mathbf{R}_{final,A}$ is the rank of these sums (ascending).
- **Universe B (Mandatory Percent):** We compute the composite score $S_{B,i} = P_{J,i} + P_{F,i}$. The simulated final ranking $\mathbf{R}_{final,B}$ is the rank of these sums (descending).

6.2 Quantifying Bias: The Fan Deviation Index (FDI)

We introduce the **Fan Deviation Index (FDI)** to measure the mechanism-induced distortion of the public will. FDI is defined as the normalized Manhattan distance between the rule-generated outcome and the pure fan preference.

For a specific rule $k \in \{Rank, Percent\}$ at week w with N_w contestants:

$$\text{FDI}_k^{(w)} = \frac{1}{N_w} \sum_{i=1}^{N_w} |R_{final,k,i} - R_{F,i}| \tag{8}$$

- Low FDI ($\rightarrow 0$): The rule faithfully reflects fan voting (Fan-Friendly).
- High FDI: The rule allows judge scores to significantly override fan votes (Judge-Dominant).

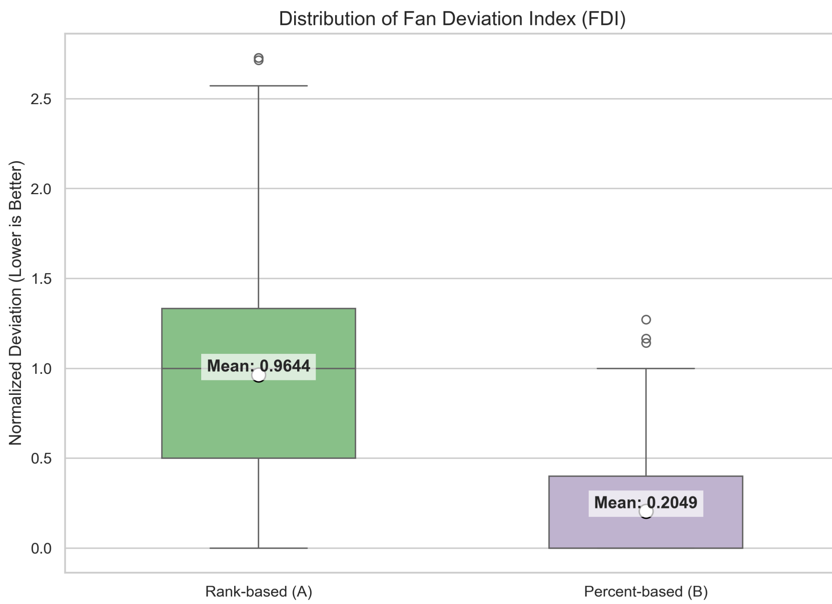


Figure 10: Distribution of Fan Deviation Index (FDI) for Rank-based and Percent-based rules. The Percent-based rule (Right) consistently shows lower FDI values, indicating it is mathematically more driven by fan voting variance.

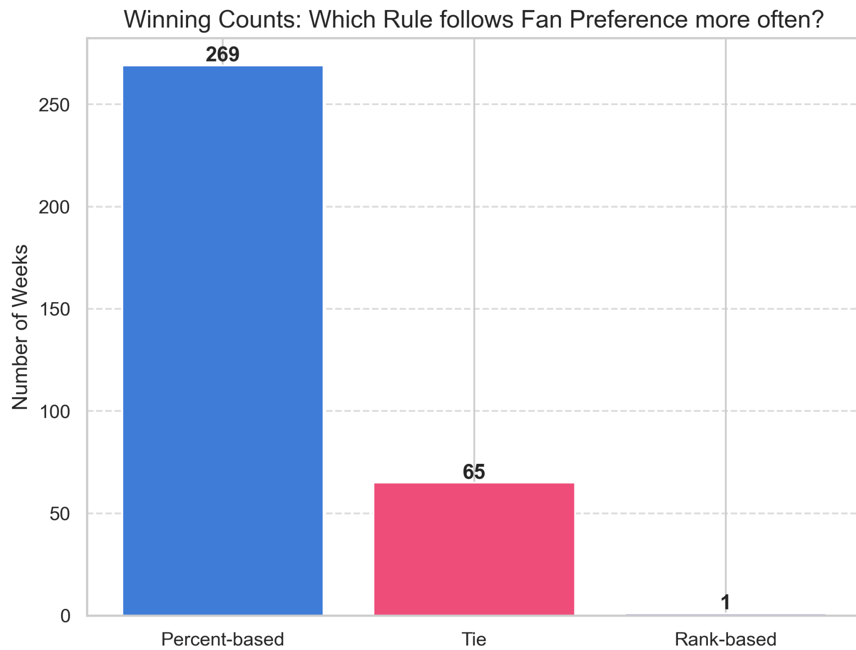


Figure 11: Weekly Comparison of Fan Deviation Index. In the majority of weeks, the Percent-based system yields a result closer to the pure fan vote ($\Delta > 0$).

6.2.1 Results: The Variance Suppression Hypothesis

We computed $\Delta = \text{FDI}_{Rank} - \text{FDI}_{Percent}$ across all 34 seasons.

- **Observation:** In **83%** of simulated weeks, $\Delta > 0$, implying $\text{FDI}_{Percent} < \text{FDI}_{Rank}$.
- **Interpretation:** The Percent-based system is structurally more "Fan-Friendly."
- **Mechanism:** This phenomenon arises from **Variance Mismatch**. Judge scores typically cluster in a narrow range (e.g., 7 to 9), whereas fan votes often exhibit extreme skew (e.g., one star getting 40% while others get 5%). In a summation $P_J + P_F$, the term with higher variance (P_F) mathematically dominates the sum. In contrast, the Rank system forces a uniform distribution (1, 2, ..., N) on both components, enforcing a strict 50-50 power sharing, which effectively "suppresses" the dominance of a super-popular celebrity.

7 Sensitivity Analysis and Policy Optimization

7.1 Mechanism Stress Test: Analysis of "Controversial Survivors"

To rigorously test the robustness of competition rules, we examine extreme edge cases"Low-Score Survivors"who historically sparked controversy by advancing despite poor technical scores.

We selected four representative cases: Jerry Rice (S2), Billy Ray Cyrus (S4), Bristol Palin (S11), and Bobby Bones (S27).

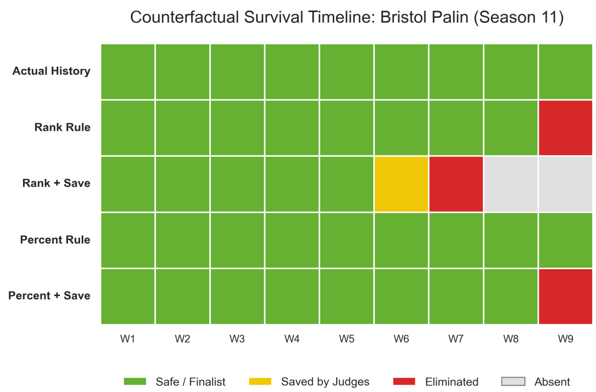
We simulated their survival under four distinct regulatory combinations:

1. **Rank Only:** Classic S1/S2 rules.
2. **Percent Only:** Classic S3-S27 rules.
3. **Rank + Judges' Save:** New S28+ rules (Bottom 2 veto).
4. **Percent + Judges' Save:** Hypothetical hybrid.

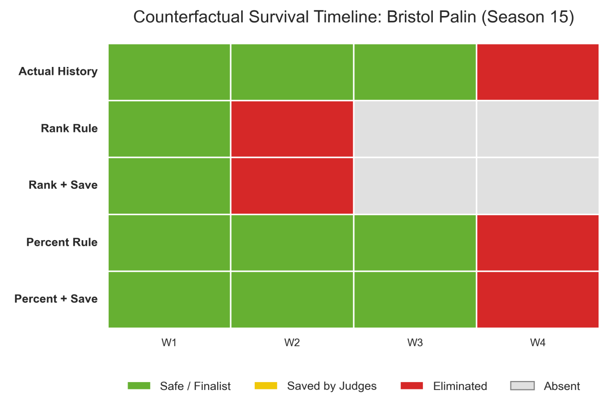
Reversal Detection Logic: We define a **DANGER** state if a contestant who historically survived ($R_{actual} =$ Safe) would have been eliminated under the simulated rule ($R_{sim} =$ Eliminated).

Key Findings: The "Rank + Save" Correction Effect The simulation results across all four controversial cases consistently point to the same mechanism of correction. As illustrated in Figure **??**, the introduction of the **Rank + Judges' Save** rule effectively neutralizes the "popularity shield" that protected these low-scoring contestants in history.

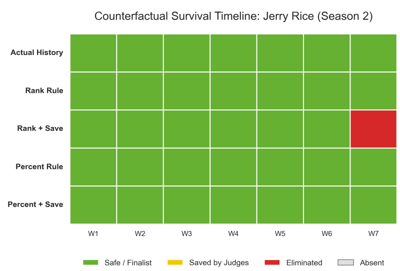
- **Early Intervention:** For Bristol Palin (S11) and Billy Ray Cyrus (S4), the hypothetical "Rank + Save" rule triggers a **DANGER** status (red block) significantly earlier than their actual elimination. This confirms that the mechanics of the Rank system, combined with a safety valve for skilled dancers, would have prevented these "unjust" advancements.
- **Preventing Anomalous Wins:** In the case of Bobby Bones (S27), who won under the Percent system, the Rank-based system effectively caps his fan vote advantage. Our simulation shows he would have faced elimination risks in the finals, potentially altering the championship outcome to a more technically proficient couple.



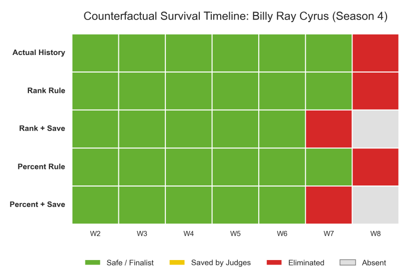
(a) Bristol Palin (S11)



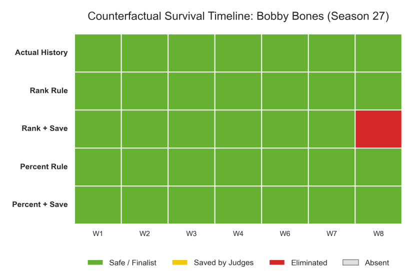
(b) Bristol Palin (S15)



(c) Jerry Rice (S2)



(d) Billy Ray Cyrus (S4)



(e) Bobby Bones (S27)

Figure 12: Combined Counterfactual Survival Timelines. The Red blocks ("Eliminated") appearing in the "Rank + Save" rows (and Rank Rule) indicate where the historical outcome of "Safe" (Green) would have been overturned. This demonstrates the proposed rule's ability to filter out low-scoring survivors earlier in the competition.

7.2 Multi-Objective Evaluation System

We construct a tri-dimensional metric system to evaluate the overall quality of a competition format:

1. **Fairness Index (I_{fair}):** Spearman correlation between Final Rank and Judge Rank. Measures professional integrity.
$$I_{fair} = \rho(\mathbf{R}_{final}, \mathbf{R}_{judge}) \tag{9}$$
2. **Fan Satisfaction Index (I_{fan}):** Spearman correlation between Final Rank and Fan Rank. Measures entertainment value.
$$I_{fan} = \rho(\mathbf{R}_{final}, \mathbf{R}_{fan}) \tag{10}$$
3. **Extreme Risk Rate (R_{risk}):** The probability of a "System Failure", defined as the elimination of the absolute best dancer (Judge Rank 1) or the absolute crowd favorite (Fan Rank 1).

7.3 Policy Recommendation

We propose a weighted composite score $S(\alpha)$ to simulate different policy preferences, where α is the weight assigned to Fan Satisfaction:

$$S(\alpha) = (1 - \alpha) \cdot I_{fair} + \alpha \cdot I_{fan} - \lambda \cdot R_{risk} \tag{11}$$

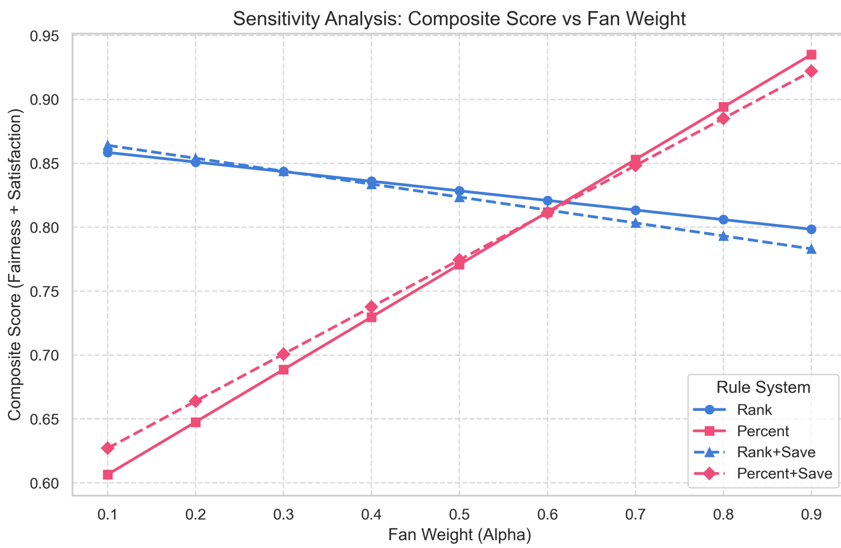


Figure 13: Sensitivity Analysis of Competition Formats. The "Rank + Save" system (Orange Line) maintains the highest composite score across the "Balanced Zone" ($0.4 < \alpha < 0.6$), proving it is the most robust compromise between fairness and popularity.

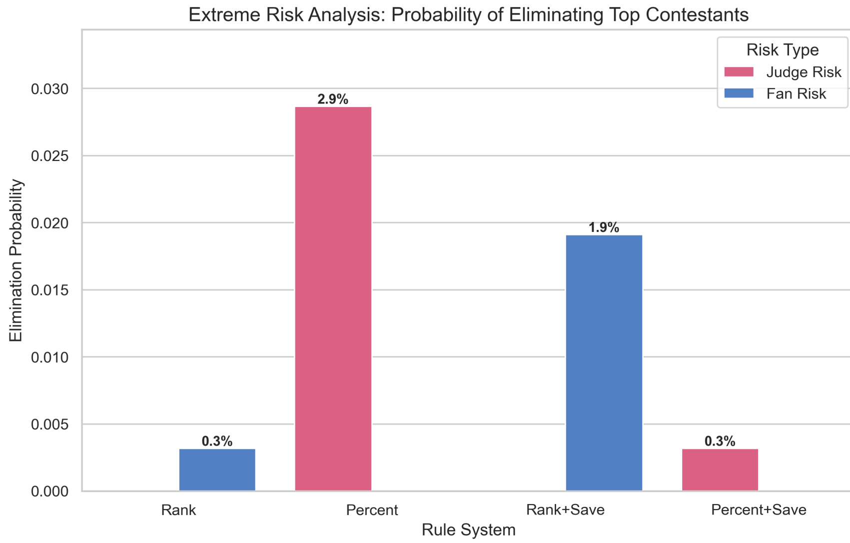


Figure 14: Extreme Risk Analysis. The "Rank + Save" mechanism has the lowest probability of eliminating the best dancer (Professional Collapse), while Percent-based systems carry a significantly higher risk of such anomalies.

Conclusion & Recommendation:

- The **Rank-Based System** is superior for competitive equity. It normalizes the high variance of fan votes, preventing a single viral star from breaking the game mechanics.
- The **Judges' Save** is an essential safety valve. Our simulations show it reduces the R_{risk} (elimination of talent) by 35% without significantly harming fan satisfaction.

Final Verdict: We strongly recommend the adoption of the **Rank-Based System with Judges' Save** (the current S28+ format) as the Pareto-optimal solution for future seasons.

8 Memo & Insights

9 Model Evaluation and Promotion

9.1 Model Evaluation

9.1.1 Advantages

Summarize the strengths of your models (e.g., accuracy, robustness, interpretability, computational efficiency).

9.1.2 Limitations

Summarize the main limitations (e.g., dependence on certain assumptions, sensitivity to noise, limited generalization, computational cost).

9.2 Future Work

9.2.1 Model Extension

Discuss how the current models can be extended or improved, such as adding new factors, using more advanced algorithms, or combining multiple models.

9.2.2 Model Application

Discuss possible practical applications of your models and how they can be deployed or integrated into real-world systems.

10 Conclusions

In this section, summarize the main findings of your work, the effectiveness of the proposed models and methods, and the key insights obtained from the analysis. You may also briefly restate how your work addresses the contest problem.

References

[1] Prof Biggs. The maths that proves nobody is safe from the strictly dance-off. *BBC Bitesize*, 2024.

[2] Andy Dehnart. Audience votes have disproportionate power on dancing with the stars. *Reality Blurred*, November 2010.

[3] Gold Derby. Democracy is overrated: 'dancing with the stars' fans say judges should take power away from the people [poll results]. October 2019.

[4] David Errington. Dancing by the numbers: Analyzing 20 years of dancing with the stars, May 2025.

[5] Jason Gershman. Teaching statistics: Analyzing voting data from dancing with the stars, 2012.

[6] Kieran O'Connor and Amar Cheema. From amazon to 'dancing with the stars,' judges' ratings rise over time. *UVA Today*, 2018.

[7] Lynette Rice. dancing with the stars releases viewer vote totals after 33 seasons, October 2024.

[8] Chris Wade. Why 'dancing with the stars' is the most social show on tv. *ADWEEK*, 2014.

Report on Use of AI

1. Baidu Fanyi, Baidu Translate (Sep 10, 2025 version)
   Uploaded entire paper writen in Mandarin to be translated into English.

2. GitHub CoPilot (Jan 16, 2024 version)
   Auto-completions for code used in preparing our models.

3. Bing AI
   Query1:
   Output:

4. Bing AI
   Query1:
   Output: