Problem Chosen	2026 MCM/ICM Summary Sheet	Team Control Number
C		2621213

Dancing with Fairness: Competition Formats Diagnostics and Optimization

Summary

Dancing with the Stars is a globally popular televised competition combining professional judging with public voting. Nevertheless, this dual evaluation system entails potential fairness issues, and the absence of publicly available fan voting data further limits transparency.

In Task 1, we developed a **multi-stage fan voting share inference model (MFSIM)** to estimate undisclosed fan voting share. Using elimination outcomes and sociological assumptions, we built a **multi-constraint dynamic Bayesian network model**. **Constrained Monte Carlo** was applied to filter feasible samples, followed by a **critical sampling technique** to compute the fan voting share. Model performance was evaluated using **consistency metrics** and **certainty metrics**. The model achieved a high $OACC$ of **93.1%** and a declining average CV from **0.317 to 0.182** as the competition progressed, demonstrating strong inferential reliability.

In Task 2, we simulated all contestants under rank-based and percent-based formats, identifying **41 inconsistent** cases. We then defined **fan power index (FPI)** to quantify fan influence. Through mathematical derivation and real competitions with high FPI and low variance, we concluded that the **percent-based format favors fan voting**. Next, we introduced **mean controversy index (MCI)** to measure contestant divisiveness, and the top 20 controversial contestants were selected. Simulations of eight high-MIC contestants across four formats revealed notable discrepancies with actual results, while the introduction of **Judge Save** effectively constrained controversial contestants. Finally, **Monte Carlo sensitivity analysis** showed that the **Rank + Judge Save** format exhibits the strongest robustness, which is recommended in the future.

In Task 3, we developed a **multi-factor impact assessment model (MFIAM)** to evaluate the effects of contestants' personal backgrounds and professional partners. After transforming individual data into statistically meaningful features, we fitted a **dual-path linear mixed-effects model**, decomposing advancement decisions into a **Judge-Led Path** and a **Fan-Led Path**, and then integrated the predicted values to explain final rankings. Results indicate that contestant background accounts for **36.13%** of outcome variance, with age **negatively** affecting both judges' scores **(-0.53)** and fan votes **(-0.35)**. Regional and industry attributes exhibit **clear divergence** across evaluation paths. Professional partners explain **12.59%** of outcome variabilty, with elite dancers such as *Derek Hough* and *Mark Ballas* providing consistent **positive** gains across both paths.

In Task 4, we developed the **log-normal Z-score fusion (LNZF) system**. We first constructed a **scoring framework** by applying **Z-score normalization** and aggregating them using a **Sigmoid function** weighting scheme. We integrated this system with the Judge Save mechanism, finding that it performs better in **selecting high-quality** contestants and **eliminating controversial** ones. To ensure rule simplicity for a televised show, we approximated the dynamic weights with fixed values - **0.6 for Weeks 1–5, 0.5 from Week 6 to the pre-final, and 0.4 in the final** - achieving a **95.44% consistency** with the Sigmoid-based rule, which demonstrates the validity of the system. Finally, we conducted a **valuation** of our model and prepared a memo for *Dancing with the Stars*.

Keywords: Dynamic Bayesian Network, Constrained Monte Carlo, Linear Mixed-Effects Model

Contents

1. Introduction . . . 3
1.1 Background . . . 3
1.2 Problem Restatement . . . 3
2. Notations & Assumptions . . . 4
3. Data Processing . . . 5
3.1 Variable Extraction and Structural Transformation . . . 5
3.2 Data Cleaning and Outlier Removal . . . 5
4. Task 1: Multi-Stage Fan Voting Share Inference Model (MSFIM) . . . 6
4.1 Multi-Constraint Dynamic Bayesian Network Model . . . 6
4.2 Constrained Monte Carlo . . . 7
4.3 Results Analysis: Dynamic Evolution of Fan Voting Share . . . 8
4.4 Model Evaluation System . . . 9
5. Task 2: Competition Formats Diagnostic and Evaluation . . . 12
5.1 Simulation Diagnosis of Competition Formats Bias . . . 12
5.2 Cases Study: Critical Review of Controversial Contestants . . . 15
5.3 Monte Carlo Sensitivity Analysis . . . 16
6. Task 3: Multi-Factor Impact Assessment Model (MFIAM) . . . 17
6.1 Dual-Path Linear Mixed-Effects Model . . . 17
6.2 Analysis of Fitting Results . . . 19
7. Task 4: Log-Normal Z-Score Fusion System (LNZF System) . . . 22
7.1 Scoring Framework . . . 22
7.2 Monte Carlo Performance Analysis and Sensitivity Analysis . . . 23
7.3 From Adaptive to Piecewise . . . 23
8. Strengths and Weaknesses . . . 23
8.1 Strengths . . . 23
8.2 Weaknesses . . . 24
References . . . 24

1 Introduction

1.1 Background

Dancing with the Stars (DWTS), a prominent American television reality show, utilizes a core mechanism that combines expert judge scores with fan popularity votes to evaluate the competitive performances of celebrity-professional pairings. This system is designed to balance **technical dance proficiency** with the **social appeal** of the contestants.

However, the program's regulations have evolved dynamically in response to seasonal controversies: the advancement of *Jerry Rice* to the finals in Season 2 despite significantly low judge scores prompted a shift from **rank-based** to **percent-based** scoring. Similarly, the controversial victory of *Bobby Bones* in Season 27 led directly to the reform in Season 28, which reintroduced the ranking method and implemented a compensatory mechanism **allowing judges to save an eliminated contestant from the bottom two**. Such rule iterations not only reflect the producers' trade-off regarding competitive fairness but also provide a research opportunity to construct more robust mathematical evaluation models through inverse modeling of latent variables. Nevertheless, the specific figures of fan votes are regarded as a closely guarded secret, making the analysis of the weight interplay between judges and the public across different aggregation methods the central challenge of this modeling task.

1.2 Problem Restatement

To thoroughly investigate the fairness and evolution of the *Dancing with the Stars (DWTS)* evaluation mechanism, this project decomposes the original problem into four core tasks and conducts a systematic analysis of data spanning 34 seasons by establishing mathematical models.

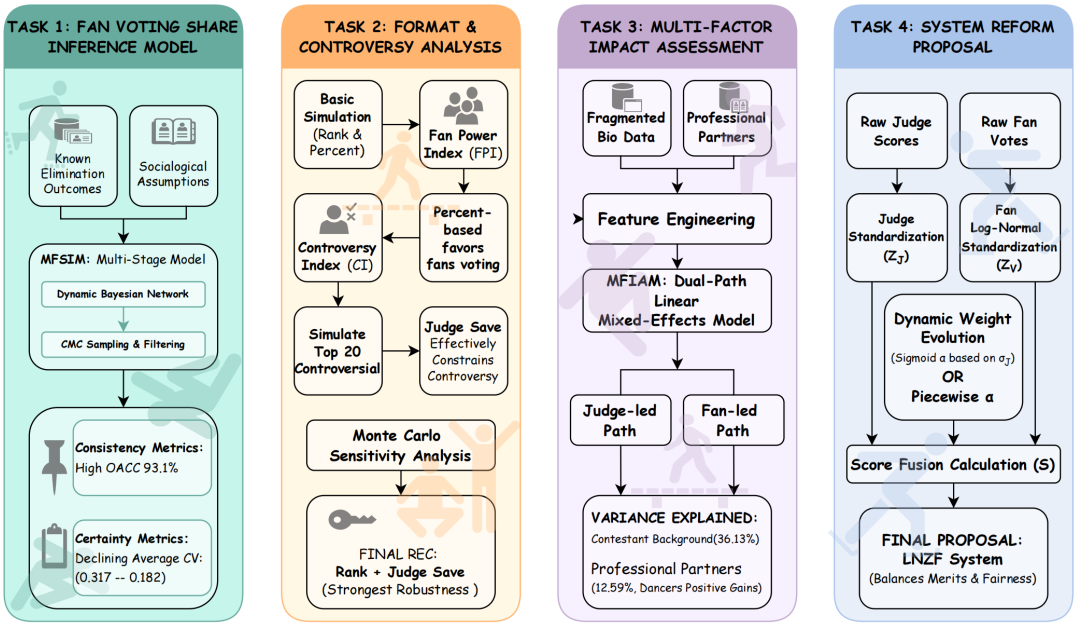

Figure 1: Our Work

Task 1: Develop an inverse estimation model to infer undisclosed fan votes and quantify the certainty and consistency of these estimates with historical outcomes.

Task 2: Compare the *Rank* and *Percent* methods regarding their balance between judges and fans. Conduct counterfactual simulations on controversial cases to assess how rule changes would alter contestants' ultimate fates.

Task 3: Investigate the influence of contestant attributes and professional partners on performance, analyzing the similar or divergent impacts of these features on judge scores versus fan votes.

Task 4: Propose an optimized evaluation framework designed to enhance a specific strategic dimension - such as fairness or audience engagement - providing data-driven justifications and strategic recommendations for producers.

2 Notations & Assumptions

Some important mathematical notations used in this paper are listed in Table 1.

Table 1: Notations used in this paper

Symbol	Description
$S_{t,i}$	judge's total score for contestant i in week t
$V_{t,i}$	fan votes share for contestant i in week t
S_{RM}	score in rank-based mechanism
S_{PM}	score in percent-based mechanism
λ	momentum factor
γ	rating weight coefficient
σ_j	deviation of judge scores
β	fixed effect coefficient vector
u	random effect vector vector
w_1	judge-path weight
w_2	fan-path weight
k	sensitivity coefficient of Sigmoid function

Given that practical problems often involve numerous complex factors, we first propose reasonable assumptions to simplify the model, with each assumption followed by a corresponding justification.

► **Assumption 1**: In the absence of major exogenous shocks, a contestant's fan base is assumed to remain relatively stable across consecutive weeks, and their supporters' voting preferences are similarly unlikely to shift dramatically.

▷ **Justification**: Fan loyalty is characterized by inertia, implying that the composition of a

contestant's supporter base does not change abruptly over short time horizons.

▶ **Assumption 2**: A contestant's judges' scores are assumed to be generally positively correlated with their audience favorability and latent popularity.

▷ **Justification**: While exceptional cases exist, overall, there is a positive correlation between fan appeal and skill proficiency, and the latter plays a central role in determining voting outcomes.

▶ **Assumption 3**: Withdrawals are modeled as random exogenous events, independent of weekly performance and without impact on the relative popularity of remaining contestants.

▷ **Justification**: Withdrawals are typically driven by injuries or other off-stage factors and therefore do not reflect the contestant's underlying popularity.

▶ **Assumption 4**: Despite inherent subjectivity, judges' scores largely capture the actual performance levels of the contestants.

▷ **Justification**: Judges generally have professional knowledge and experience, allowing them to provide relatively objective assessments of contestants' technical performance.

3 Data Processing

Before constructing the mathematical model, it is necessary to clean and reorganize the original historical dataset of *Dancing with the Stars (DWTS)*.

3.1 Variable Extraction and Structural Transformation

3.1.1 Judges' Score Aggregation

The processing procedure retrieves and sums all valid scores for each contestant by week and contestant index, resulting in the total weekly judges' score $S_{t,i}$ for each contestant . According to season-specific rules, these totals are then converted into **contribution proportions** to accommodate different scoring systems. Contestants who did not perform in a given week (marked as *N/A* or 0) are excluded from the weekly calculation matrix.

3.1.2 State Label Mapping

To formalize model constraints, we mapped raw `results` data into a structured system: (1) Boolean categories (*Eliminated*, *Safe*, or *Ignored*) based on weekly status; and (2) Ordinal features derived from final placements to enforce seasonal rank constraints.

3.2 Data Cleaning and Outlier Removal

3.2.1 Normalization

To standardize disparate scoring scales, judges' scores are normalized via **min-max scaling**:

$$S'_{t,i} = \frac{S_{t,i} - \min(S_t)}{\max(S_t) - \min(S_t) + \delta} \tag{1}$$

where δ prevents division by zero. This maps scores to the $[0, 1]$ interval, serving as stable priors for subsequent probabilistic inference.

3.2.2 Exclusion of Withdrawal Interference

Withdrawals, driven by stochastic off-stage factors, are decoupled from popularity or performance metrics. We exclude these cases from the elimination decision set during preprocessing.

4 Task 1: Multi-Stage Fan Voting Share Inference Model (MSFIM)

4.1 Multi-Constraint Dynamic Bayesian Network Model

4.1.1 Dynamic Priors: Momentum Propagation and Weight Assignment

Rather than treating weekly vote shares as independent events, the model frames them as an evolutionary process with strong temporal dependence. By integrating a probabilistic framework with historical behavioral inertia, this module constructs a plausible initial solution space.

(1) Dirichlet Distribution

To represent the fan votes share V as a multivariate variable on a simplex, we employ a **Dirichlet distribution** as the prior:

$$V_t \sim \text{Dir}(\alpha_t), \quad \alpha_t = (\alpha_{t,1}, \ldots, \alpha_{t,n}) \tag{2}$$

This framework naturally enforces the physical constraint $\sum V_{t,i} = 1$ while allowing efficient sampling of the competitive landscape.

(2) Momentum Propagation Mechanism

To capture behavioral inertia within fan communities, the Dirichlet parameters are updated via a **time-dependent** equation that anchors vote share evolution:

$$\alpha_{t,i} = \alpha_{base} + \gamma \cdot S'_{t,i} + \lambda \cdot \hat{V}_{t-1,i} \tag{3}$$

where the **momentum factor** λ ensures longitudinal continuity. This mechanism enables the model to refine its estimation of a contestant's latent popularity based on cumulative historical performance.

4.1.2 Likelihood Constraints: Binary Logic-Based Search Space Pruning

The inference process employs a **binary likelihood function** that partitions the sampling space into feasible and infeasible regions via logic-based cutoffs. For any simulated sample $\mathbb{V}_{sim}$, the likelihood is defined as:

$$L(V_{sim} \mid \text{Data}_t) = \begin{cases} 1, & \text{if } \Phi(V_{sim}, J_t, \text{Rules}) \text{ is consistent} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

This hard constraint ensures that only vote-share distributions capable of reproducing the observed elimination outcomes are retained for posterior estimation.

4.1.3 Competition Rule Constraints: Cross-Season Adaptive Decision Criteria

Serving as the basis for the likelihood function, the model automatically loads the corresponding set of competition rule constraints $\mathbb{C}$ according to the season index $\mathbb{S}$. This set of constraints determines the decision logic of the Φ function:

(1) Rank-Based Logic (S1–S2)

Under the rank-based competition rules in the early stages of the show, the model requires that samples satisfy the condition that *the sum of an eliminated contestant's judges' ranking and audience vote ranking must be the highest among all contestants*. Its mathematical expression is:

$$S_{RM} = \text{Rank}(V_{sim}) + \text{Rank}(J) \qquad S_{RM,e} \geq \max_{s \in S_a J_e} S_{RM,s} + \epsilon \tag{5}$$

(2) Percent-Based Logic (S3-S27)

In the over-20-season percent-based competition format, the decision criterion shifts to requiring that the overall percentages generated by the samples satisfy the condition that *eliminated contestants must have lower percentages than those who survive*. Its mathematical expression is:

$$S_{PM} = P_{\text{judge}} + V_{\text{fan}} \qquad S_{PM,e} \leq \min_{s \in S_a J_e} S_{PM,s} + \epsilon \tag{6}$$

(3) Judge's Save Logic (S28-S34)

The model requires that the inferred popularity V_{sim} ensures that eliminated contestants *fall within the Bottom 2 zone* and *cannot be preferentially saved* given the judges' weighting preferences. Its mathematical expression is:

$$S_e \leq \text{Second\_Lowest}(S_{all}) \qquad W_{\text{judge},e} + \text{Error} < W_{\text{judge},s} \tag{7}$$

4.1.4 Smoothing Constraint: Asymmetric Inertia in Popularity Evolution

To align with the sociological principle of **popularity inertia**, the model imposes an asymmetric smoothing constraint on temporal vote-share fluctuations:

$$\text{Filter: } -3.5\% \leq V_{t,i} - V_{t-1,i} \leq 25\% \tag{8}$$

This asymmetric smoothing window captures the fanbase's psychological tendency that *attrition is sticky while growth is explosive*, further enhancing the model's robustness in handling anomalous elimination events.

4.2 Constrained Monte Carlo

Given the nonlinear rules and binary likelihood, the posterior distribution lacks a closed-form analytical solution. Consequently, we employ a **constrained Monte Carlo (CMC)** algorithm to approximate the true vote shares via iterative sampling within a high-dimensional simplex.

4.2.1 Random Sampling and Filtering in the Constrained Space

The core of the algorithm lies in approximating the posterior of the latent vote distribution through large-scale random sampling.

(1) Simplex Sampling

At each week t, the algorithm draws $N = 20,000$ candidate vote-share vectors from a **Dirichlet distribution** parameterized by the dynamically constructed prior α_t.

(2) Hierarchical Filters

Each candidate sample must sequentially satisfy the **smoothing constraint** and the **competition rule constraint**. The former ensures plausible temporal evolution of vote shares, while the latter enforces logical consistency with observed elimination outcomes.

4.2.2 Critical Sample Drawing and Posterior Parameter Estimation

After obtaining the set of valid samples V_{valid}, the algorithm does not simply take the mean; instead, it employs a **critical sampling** technique.

(1) Critical-Point Locking for Eliminated Contestants

For contestants eliminated in the current week, their vote shares should theoretically lie at the highest levels *sufficient to cause their elimination*. Therefore, the algorithm employs a **maximum likelihood boundary search logic** to extract the critical samples from the set in which the contestant's vote share falls within the top 30%.

$$\hat{V}_{t,\text{elim}} = \mathbb{E}\left[V \mid V \in \text{Top 30\% valid samples}\right] \tag{9}$$

(2) Robust Moment Estimation for Safe Contestants

For contestants who safely advance, their vote-share distributions are generally wider and not directly constrained by the elimination threshold. The model computes the final vote-share estimate $\hat{V}_{t,i}$ by robustly averaging over the aforementioned critical sample set.

$$\hat{V}_{t,\text{safe}} = \frac{1}{M} \sum_{m=1}^{M} V_{t,i}^{(m)} \tag{10}$$

4.2.3 Logical Loop Backtesting Verification

As a final validation, the mean estimates $\hat{V}_{t,i}$ are backtested against the competition rule equations. If the resulting elimination list matches historical records exactly, the model state is marked as **Verified**.

4.3 Results Analysis: Dynamic Evolution of Fan Voting Share

After running the Monte Carlo simulations with a fixed seed (SEED=2026) for full reproducibility, we extract the dynamic estimates of **fan voting shares**. To illustrate competitive dynamics across seasons, we plot each contestant's estimated vote share per week along with corresponding standard deviation bands.

To examine the underlying dynamics across competition stages, we select two representative cases from each of the two core periods, S3–S28 and S29–S34, from the 34 trend plots. These cases capture typical vote trajectories at elimination and reveal characteristic patterns of fan voting under different mechanisms, providing clear examples for understanding the intrinsic logic of eliminations across formats.

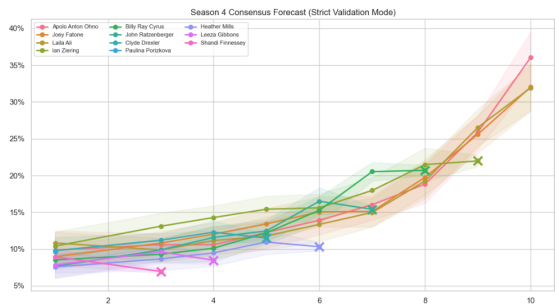

Figure 2: Season 4

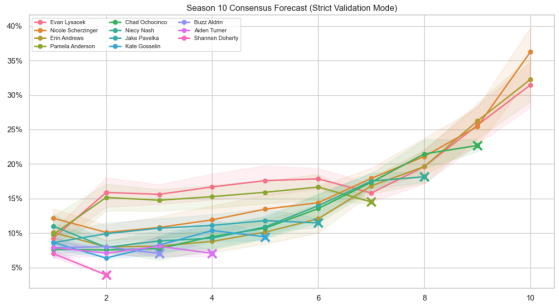

Figure 3: Season 10

(1) Percent-Based Voting: Relative Growth Competition

Under the percent-based logic from Seasons 3 to 27, the share of votes released by eliminated contestants is redistributed among the remaining participants, resulting in a generally upward trend in vote shares. In this context, elimination is determined less by the absolute loss of supporters and more by lagging relative growth. If a contestant's percentage increase is slower than that of their competitors, their proportion within the total vote pool is diluted, making it difficult to keep pace with the explosive expansion of the vote pool and bringing them closer to the elimination threshold.

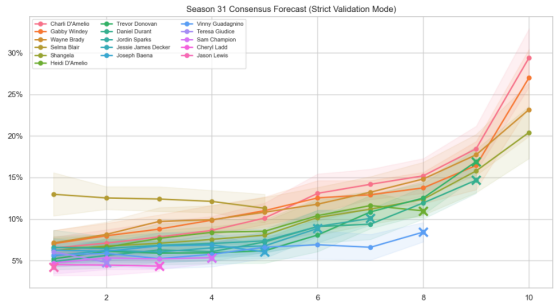

Figure 4: Season 31

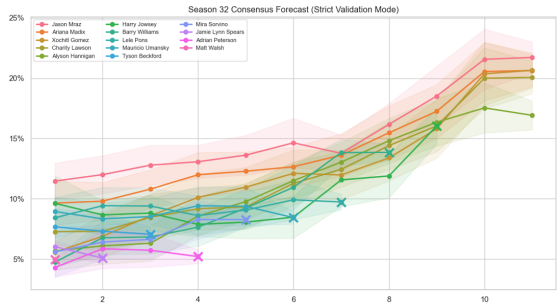

Figure 5: Season 32

(2) Rank-Based Voting: Survival Space Competition

From Seasons 1 to 2, and 29 to 34, the rank-based system led to a more nonlinear and compressed pattern of eliminations, especially after the introduction of judge save mechanism. Under this framework, the influence of judges increases relatively, so that even if the vote share of lower- and mid-tier contestants grows with the release of additional votes, it often remains confined within a narrow percentage range, making it difficult to surpass the elimination threshold. Although judge saves occur after voting, they determine the continuity of vote share trajectories: contestants with superior technical performance can survive despite low vote shares.

4.4 Model Evaluation System

4.4.1 Consistency Measurement: Logical Self-Consistency Verification

The consistency metric C_t is a binary indicator used to assess whether the inferred posterior mean $\hat{V}_t$ can reproduce the actual elimination outcomes under the preset competition rules $\mathbb{C}$.

$$C_t = \mathbb{I}\left[\text{FormatLogic}\left(\hat{V}_t, S_{judge,t}\right) \equiv \text{Observed\_Result}_t\right] \tag{11}$$

If $C_t = 1$, it indicates that the popularity distribution for that week is mathematically and logically consistent. If $C_t = 0$, it reveals a potential *logical break* in that week. Therefore,the **season accuracy** with N elimination weeks ACC_{season} and **overall accuracy** $OACC$ can be expressed as:

$$ACC_{\text{season}} = \frac{1}{N}\sum_{t=1}^{N} C_t \qquad OACC = \frac{\sum_{s=1}^{S} ACC_{season}}{\sum_{s=1}^{S} N_s} \tag{12}$$

The model exhibits high logical consistency, achieving an **overall accuracy of 93.1%**. During the percent-based phase, accuracy remained near **100%**, reflecting the deterministic nature of share-allocation logic. Although the transition to rank-based systems and judge interventions introduced minor fluctuations, perfect predictions in seasons like Season 31 underscore the model's precision in capturing complex, modern contest dynamics.

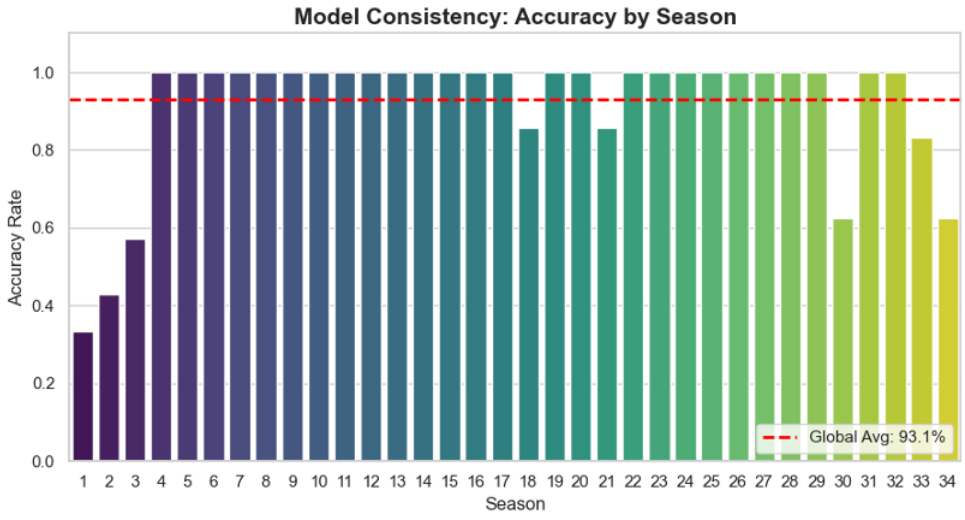

Figure 6: Season Accuracy and Overall Accuracy

4.4.2 Certainty Quantification: In-depth Analysis of Certainty

Our Bayesian inference framework not only generates point estimates of fan vote shares but, more importantly, provides complete posterior probability distributions. This allows us to rigorously quantify the certainty of each estimate. Through two complementary metrics—the coefficient of **variation (CV)** and the **confidence interval width (CIW)**—we construct a multi-dimensional certainty evaluation system that reveals profound connections between the model's inherent inference characteristics and the dynamics of the competition.

(1) Relative Quantification via Coefficient of Variation

A systematic analysis of all 6,142 contestant-week observations reveals the statistical characteristics of the model's certainty distribution. The coefficient of variation (CV), defined as:

$$\text{CV}_i = \frac{\sigma_i}{\mu_i} \tag{13}$$

where μ_i represents the posterior mean vote share and σ_i denotes the posterior standard deviation for contestant i. The model exhibits moderate inference confidence with a median CV of 0.2821. The distribution is typically right-skewed, with the 25th, 75th, and 90th percentiles at 0.2109, 0.3470, and 0.4097, respectively. This profile confirms that while the framework maintains high certainty for the majority of estimates, it retains sufficient flexibility to accommodate high-entropy edge cases.

Particularly noteworthy are the extreme cases with the highest uncertainty. Table 3 lists the five predictions with the largest CV values, all of which occur in the first week of their respective seasons. These high-uncertainty cases are not model defects but accurately reflect the genuine ambiguity present in the early stages of competition when information is scarce.

Table 2: Top Five Most Uncertain Prediction Cases

Contestant	Season	Week	Vote Estimate	CV
Sadie Robertson	19	1	0.057406	0.905642
Jen Affleck	34	1	0.063470	0.849064
Sherri Sheperd	14	1	0.068632	0.803535
Lele Pons	32	1	0.050176	0.744952
Baron Davis	34	1	0.065774	0.738516

Certainty improves systematically as the model accumulates historical data. Between Week 1 and Week 11, the average CV dropped from 0.317 to 0.182—a 43% reduction in relative uncertainty. This convergence validates the momentum propagation mechanism, where the 15x weighting of prior voting results creates an increasingly robust informational foundation for late-season predictions.

Competition rules also influence certainty levels. Percent-based seasons (S4–S27) yield a lower average CV of 0.267 due to their direct score-to-vote mapping. Conversely, rank-based seasons with judge-save mechanisms (S28–S34) exhibit a higher CV of 0.341, as subjective interventions introduce systemic complexity that heightens inferential uncertainty.

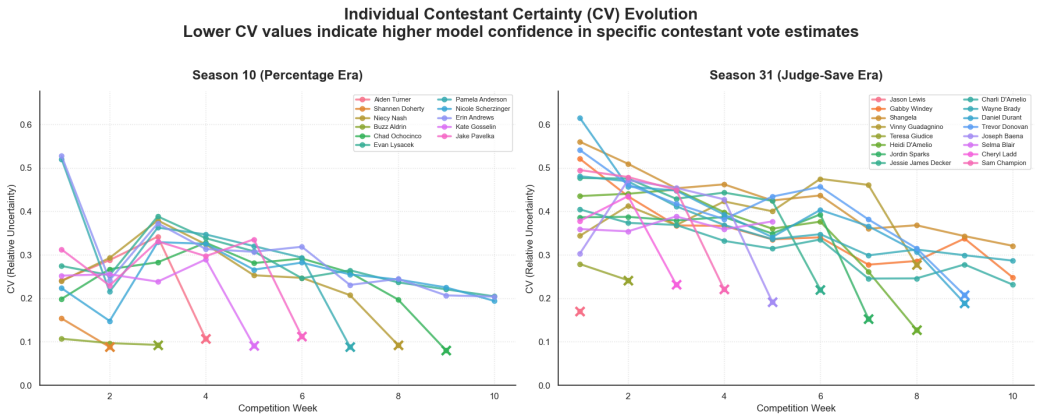

Figure 7: CV Trends in Contestant Vote Predictions

(2) Absolute Precision Assessment via Confidence Intervals

To assess absolute precision, we define the 95% confidence interval as $\text{CI}_i = \hat{\mu}_i \pm 1.96\sigma_i$. Across all observations, the average **confidence interval width (CIW)** is 12.18%, placing the true voting share typically within $\pm 6.09\%$ of the point estimate. The precision distribution categorizes 31.2% of estimates as high precision (CIW $< 3\%$), 42.7% as moderate ($3\% \leq$ CIW $< 6\%$), and 26.1% as lower precision (CIW $\geq 6\%$), reflecting a highly concentrated confidence profile.

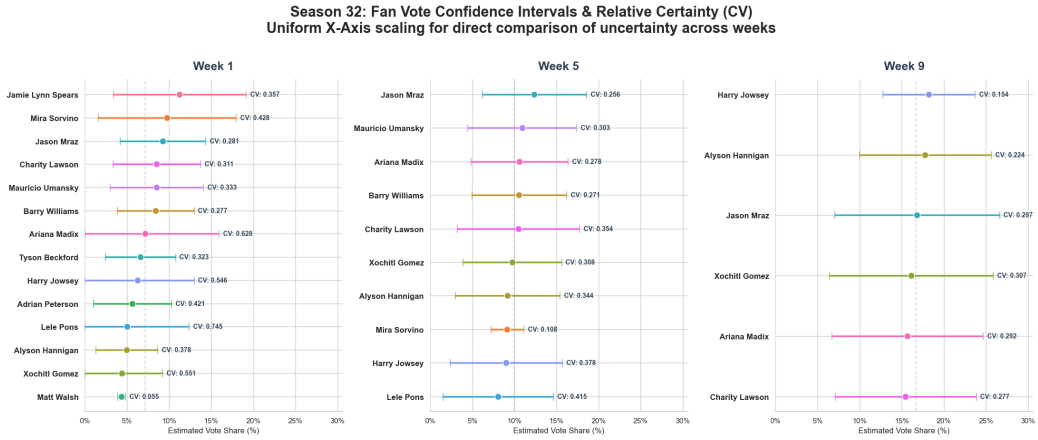

Figure 8: CI Trends in Contestant Vote Predictions

The paradoxical widening of confidence intervals in later stages reflects increased statistical sensitivity driven by sample size reduction and score compression. As elite contestants' scores converge, the model must encompass broader fan-vote ranges to remain logically consistent with potential rankings. However, the declining Coefficient of Variation (CV) proves that relative inference precision actually improves as it scales with higher mean vote shares.

5 Task 2: Competition Formats Diagnostic and Evaluation

5.1 Simulation Diagnosis of Competition Formats Bias

This section aims to use simulations based on real competition data to identify the differences in evaluation systems under various conditions and to reveal the mathematical triggering mechanisms that lead to shifts toward popular preferences.

5.1.1 Conflict Matrix Analysis

We run simulation procedures under both the rank-based system and the percent-based system, with 88.5% of cases where the two systems give identical elimination results. In 41 inconsistent cases, those where the percent-based system eliminates a high-scoring contestant while the rank-based system does not are marked in red, while those where the rank-based system eliminates a high-scoring contestant while the percent-based system does not are marked in yellow.

5.1.2 Mathematical Foundations of Competition Format Bias

(1) Fan Power Index (FPI)

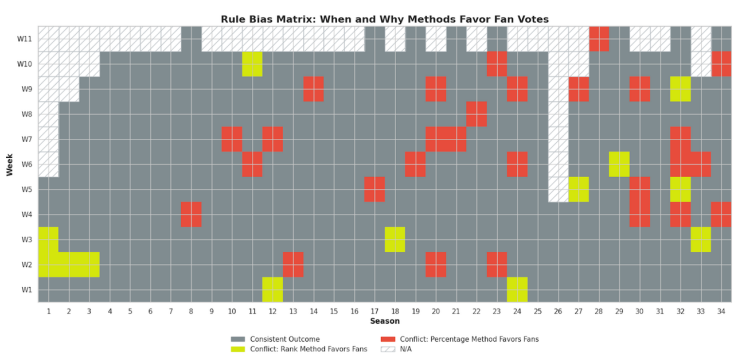

Figure 9: Conflict Situations

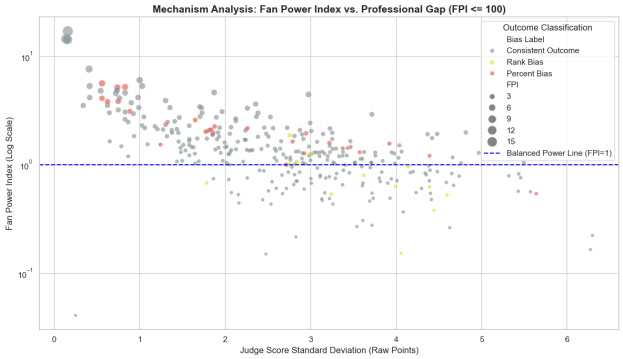

Figure 10: FPI - Variance Scatter Plot

To quantitatively measure the effective weight of fan voting under different scoring environments, we introduce the **fan power index (FPI)**. Its mathematical definition is given as follows:

$$FPI = \frac{\sigma(P_{fan})}{\sigma(P_{judge})} \tag{14}$$

This index reflects the relative strength of fluctuations in fan votes compared with fluctuations in judges' scores. When $FPI > 1$, it indicates that the noise from popular opinion overwhelms the signal of professional evaluation, implying that fan groups exert a dominant influence on the competition outcome. By plotting a scatter plot of FPI versus variance, we can derive the following key conclusions.

Conclusion 1: The fan power index exhibits an inverse correlation with the standard deviation of professional scores.

(2) Theorem 1: Fan Dominance under the Percent-Based System

Proposition: Under the percent-based system, as FPI increases, the marginal effect of FPI remain positive, and fan dominates the system.

Proof: Assuming independence between judges' scores and fan votes ($\rho = 0$), the total score variance under the percent-based system can be expressed as a linear additive form:

$$\sigma_S^2 = \sigma_{N(J)}^2 + \sigma_{N(F)}^2 \tag{15}$$

We define η as the proportion of total score variation contributed by fan voting, i.e., the explanatory share of fans in the overall variance:

$$\eta = \frac{\sigma_{N(F)}^2}{\sigma_S^2} = \frac{\sigma_{N(F)}^2}{\sigma_{N(J)}^2 + \sigma_{N(F)}^2} = \frac{FPI^2}{1 + FPI^2} \tag{16}$$

Differentiating the control equation shows that its derivative is strictly positive, and that the influence converges to 1 as $FPI \to \infty$.

$$\frac{d\eta}{d(FPI)} = \frac{2 \cdot FPI}{(1 + FPI^2)^2} > 0 \quad \text{and} \quad \lim_{FPI \to \infty} \eta = 1 \tag{17}$$

When we introduce the correlation coefficient between judges' scores and fan votes, a similar conclusion can be obtained.

$$\eta(FPI) = \frac{FPI^2 + \rho \cdot FPI}{FPI^2 + 2\rho \cdot FPI + 1} \tag{18}$$

(3) Theorem 2: Variance-Stabilizing Mechanism of the Rank-Based System

Proposition: Under the rank-based system, as FPI increases, the marginal effect of FPI remains zero, and professional judging preserves the authority.

Proof: The rank-based system maps the continuous variable J_i onto a discrete set $R_J \in 1, 2, \ldots, n$, such that its variance depends solely on the number of contestants n:

$$\sigma^2_{R(J)} = \frac{n^2 - 1}{12} = \text{Constant} \tag{19}$$

Under the rank-based system, the denominator of the effective power index FPI_{rank} is locked in as a constant and is no longer affected by compression in the original scoring scale:

$$FPI = \frac{\sigma_{R(F)}}{\sigma_{R(J)}} = \frac{\sigma_{R(F)}}{\sqrt{(n^2 - 1)/12}} \tag{20}$$

When contestants' performance levels converge, causing the variance of judges' raw scores $\sigma_{N(J)} \to 0$, the convergence behavior of the rank-based system can be expressed as:

$$\frac{d\eta}{d(FPI)} = 0 \tag{21}$$

(4) Theoretical Summary

Based on the mathematical derivations above, we can draw the following key conclusion.

Conclusion 2: Percent-based format exhibits a positive marginal effect of FPI while rank-based format nullifies this effect.

5.1.3 Empirical Distribution and Final Inference

By tracing back real competition data from 34 seasons, we can construct the distribution of judges' score variance and the panoramic FPI landscape.

Kernel density estimation indicates that the majority of samples concentrated around the median standard deviation of judges' scores, σ_J, is 2.72. The panoramic FPI landscape shows that 77.9% of samples have $FPI > 1$. Therefore, we can get the following significant fact.

Fact: The system operates in the long term within a low-variance, high-FPI landscape.

Based on the facts and conclusions above, we can infer that **in most cases, the influence of the fan on the outcome is greater under the percent-based system**. This means that even a slight reduction in performance differences can rapidly amplify bias in elimination results under the percent-based system, while the rank-based system can maintain stability through its inherent structural resistance.

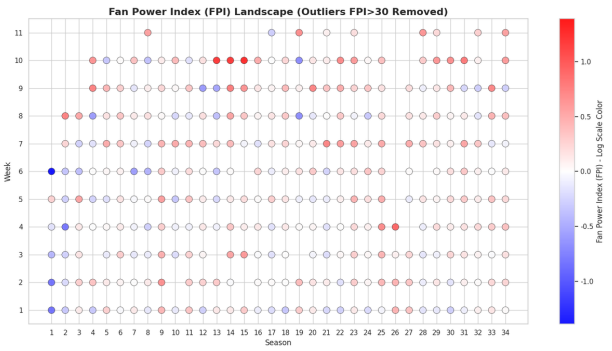

Figure 11: FPI Landscape (Remove Outliers)

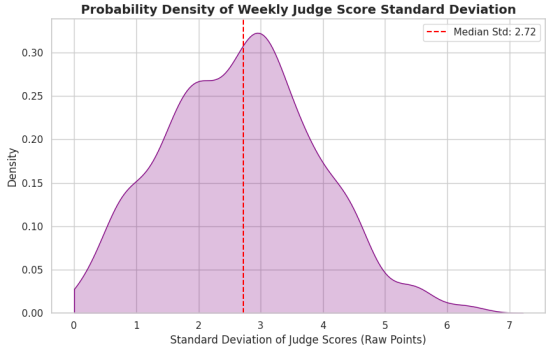

Figure 12: Variance Kernel Density Plot

5.2 Cases Study: Critical Review of Controversial Contestants

This section revisits the advancement trajectories of historically controversial contestants with the aim of evaluating the practical performance of different evaluation systems under extreme deviation cases.

5.2.1 Identification of Controversial Contestants Based on the Controversy Index

To quantitatively identify contestants exhibiting performance - popularity misalignment, we define **controversy index (CI)** and **mean controversy index (MCI)**:

$$CI_t = |\text{Rank}_{Judge,t} - \text{Rank}_{Fan,t}| \qquad MCI = \frac{1}{N}\sum_{t=1}^{N} CI_t \tag{22}$$

A CI value, with larger magnitudes, indicates that a contestant is increasingly relying on popularity to mask deficiencies in actual performance. By sorting the MCI values in descending order, we find that the examples provided in the problem are all included within the top 20 cases. This shows that the MCI is a highly effective metric for detecting controversial contestants.

Table 3: Retrospective simulation results (Examples cited in the problem are highlighted in bold)

Contestant	Season	Historical	Rank	Rank+JS	Percent	Percent+JS
Jerry Rice	**2**	**Rank 2**	**Elim W5**	**Elim W5**	**Elim W5**	**Elim W5**
Billy Ray Cyrus	**4**	**Elim W8**	**Elim W8**	**Elim W8**	**Elim W8**	**Elim W7**
Ty Murray	8	Elim W10	Elim W10	Elim W10	Elim W10	Elim W9
Bristol Palin	**11**	**Rank 3**	**Elim W6**	**Elim W5**	**Rank 3**	**Elim W7**
Bill Engvall	17	Rank 4	Rank 4	Elim W8	Rank 4	Elim W7
Terrell Owens	25	Elim W8	Elim W2	Elim W2	Elim W8	Elim W2
Bobby Bones	**27**	**Rank 1**	**Rank 3**	**Elim W8**	**Rank 3**	**Elim W8**
Xochitl	32	Rank 1	Rank 2	Rank 2	Rank 3	Rank 3

5.2.2 Simulation-Based Audit under Four Competition Formats

We constructed four competition models (*Rank*, *Rank + JS*, *Percent*, *Percent + JS*) and conducted retrospective simulations for the four given contestants and four highe-MCI contestants.

The results demonstrate that the introduction of the Judge Save mechanism is the most effective means of resolving controversy: it effectively intercepts contestants who survive solely due to overwhelming popularity. Taking *Bristol Palin* and *Bobby Bones* as representative cases, simulations show that after the incorporation of the Judge Save mechanism, both were eliminated prior to the finals, thereby safeguarding the professional standards of the competition.

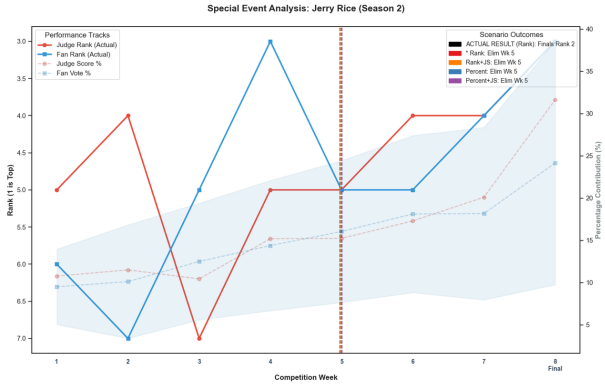

Figure 13: Jerry Rice

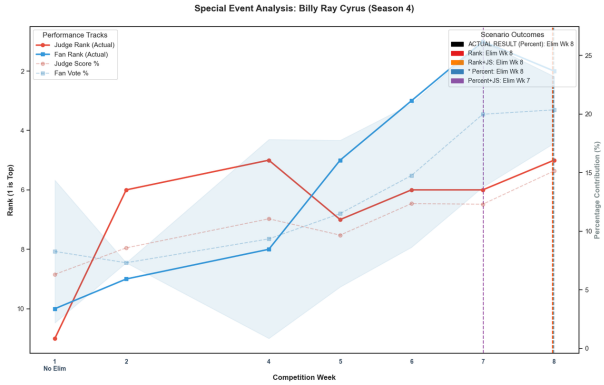

Figure 14: Billy Ray Cyrus

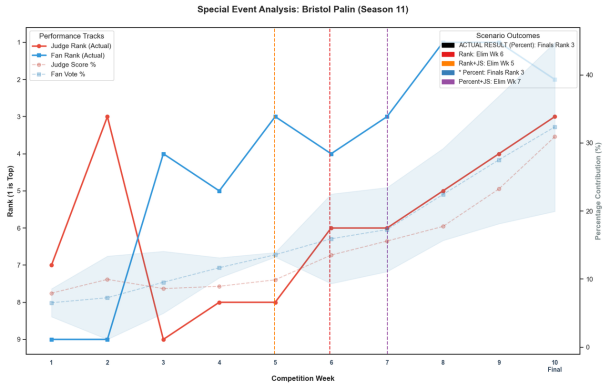

Figure 15: Bristol Palin

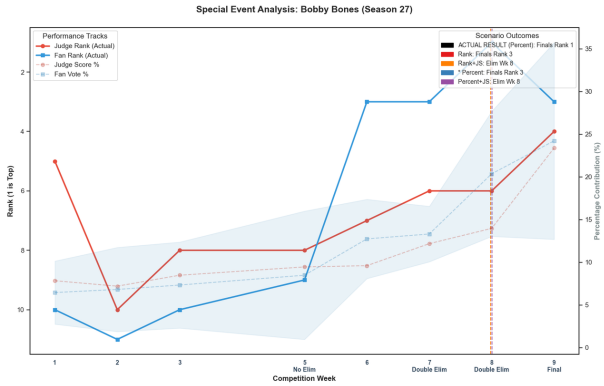

Figure 16: Bobby Bones

5.3 Monte Carlo Sensitivity Analysis

Building on the conclusions previously obtained, we further compare the robustness of the four competition formats through **Monte Carlo sensitivity analysis** to identify the optimal configuration under the current conditions.

5.3.1 Variable Selection

(1) Divergence Degree

This study selects the **divergence degree** as an independent variable, representing the strength of the negative correlation between judges' professional evaluations and fan popularity. The values range from 0 to 0.9, representing a progression from no correlation to strong negative correlation.

(2) Injustice Rate

We define the **injustice rate** as the probability that the contestant with the lowest judge score Worst_J ultimately avoids elimination in $N = 5000$ simulation runs. Its calculation formula is:

$$Y = \frac{\sum_{i=1}^{N} \mathbb{I}\left(\text{Eliminated}_i \neq \text{Worst\_J}_i\right)}{N} \tag{23}$$

5.3.2 Simulation Results and Evaluation of Competition Formats

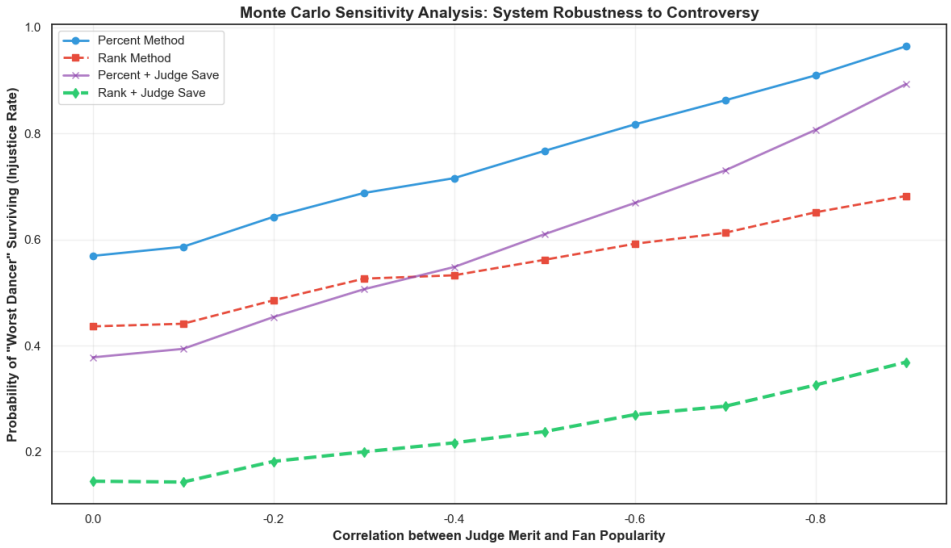

Figure 17: Monte Carlo Sensitivity Analysis Results

The Monte Carlo simulation results show that the **Rank + Judge Save** system performs the best across all simulations, with its injustice rate curve consistently remaining at the lowest level among all competition formats (including under extreme conditions), which is also consistent with the conclusions we previously obtained.

6 Task 3: Multi-Factor Impact Assessment Model (MFIAM)

6.1 Dual-Path Linear Mixed-Effects Model

To isolate intrinsic contestant traits from environmental factors and quantify divergent preferences between judges and the audience, we employ a **dual-path linear mixed-effects model**.

6.1.1 Feature Engineering

To enhance interpretability and quantify "background advantages", we utilize **feature engineering** to transform fragmented contestant data into robust, statistically significant indicators.

(1) Industry Success Probability

This feature assigns a **winning potential** score to each occupation by calculating the historical probability of industry-specific contestants reaching the top three, effectively quantifying the competitive advantage of different social backgrounds.

$$P(\text{Success})_{ind} = E\left[\mathbb{I}\left(\text{Placement} \leq 3\right) \mid \text{Industry} = \text{ind}\right] \tag{24}$$

(2) Nonlinear Transformation of Age

Accounting for the non-linear decline in physical stamina and learning speed, the model discretizes age into three stages: **young (<30)**, **middle-aged (30–50)**, and **older (>50)**. This categorization enables the capture of asymmetric scoring weights across distinct life stages.

$$\text{Age}_{bin} = \begin{cases} 1, & \text{Age} < 30 \\ 2, & 30 \leq \text{Age} \leq 50 \\ 3, & \text{Age} > 50 \end{cases} \tag{25}$$

(3) Region Clustering

The original 25 subregions are clustered into four macro-groups–*US_Native*, *Anglosphere*, *Europe_Eurasia*, and *Latin_Asia_Other*–based on geographic and cultural affinity. This aggregation increases group-level sample sizes and mitigates noise, ensuring robust identification of cross-national background effects.

$$\text{Group}_i = \sum_{j \in \text{Cluster}_i} \text{Region}_j \tag{26}$$

6.1.2 Dual-Path Learning Layer

The dual-path learning layer achieves a deep decomposition of the evaluation mechanism and variance separation by fitting two independent linear mixed-effects models.

(1) Binary Decomposition of Decision Paths

The model decomposes decision-making into a **Judge-Led Path (Technical)**, driven by professional expertise, and a **Fan-Led Path (Emotional)**, driven by audience resonance. These pathways map the independent response trajectories of judges' scores and fan votes, respectively.

(2) Environmental Decomposition of Nested Data

The sample exhibits a typical "contestant–partner" nested structure. A contestant's final performance depends not only on personal characteristics but also significantly on the professional skill of their dance partner and the partner's existing fan base. To account for this, we introduce random effects, setting a random intercept grouped by "Partner". By treating the partner as "environmental noise" and isolating its influence, the model ensures that the regression coefficients provide an unbiased estimate of the effect of the contestant's individual background on outcomes.

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{u} + \epsilon \tag{27}$$

(3) Quantifying Differences Across Dual Paths

By standardizing both judges' scores and fan votes using a StandardScaler, the model allows a direct comparison of the coefficients of the same background features across the two paths. This enables the quantification of preference shifts between professional and popular aesthetics along different dimensions.

6.1.3 Integrated Decision Layer

The integrated decision layer constructs a dynamic power-attribution framework by taking the predicted values from the previous paths ($\hat{y}_{judge}$ and $\hat{y}_{fan}$) as inputs for a secondary fitting.

$$\text{Placement} = w_0 + w_1 \cdot \hat{y}_{judge} + w_2 \cdot \hat{y}_{fan} + \epsilon \tag{28}$$

This layer quantifies the relative contributions of the two paths under different competition formats through the weights w_1 and w_2.

6.2 Analysis of Fitting Results

We analyze the fitted results of the dual-path model to quantify the influence of star contestants' background characteristics and decompose the preference differences among evaluation agents.

6.2.1 Background Effect

(1) Overall Predictive Performance

Contestants' intrinsic background explains a substantial portion of ranking variance, accounting for 36.13% of the total, indicating that innate traits largely pre-determine outcomes.

(2) Ballroom Partner Contribution

Partner assignment contributes significantly to ranking variance, explaining 12.59% of the total. Top partners like *Derek Hough* notably boost contestant rankings (median to 3rd place).

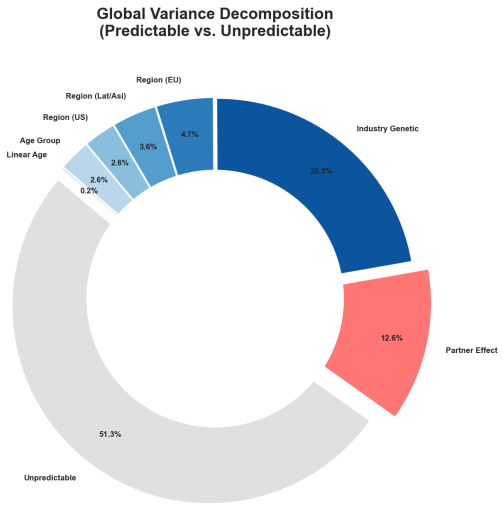

Figure 18: Variance Attribution

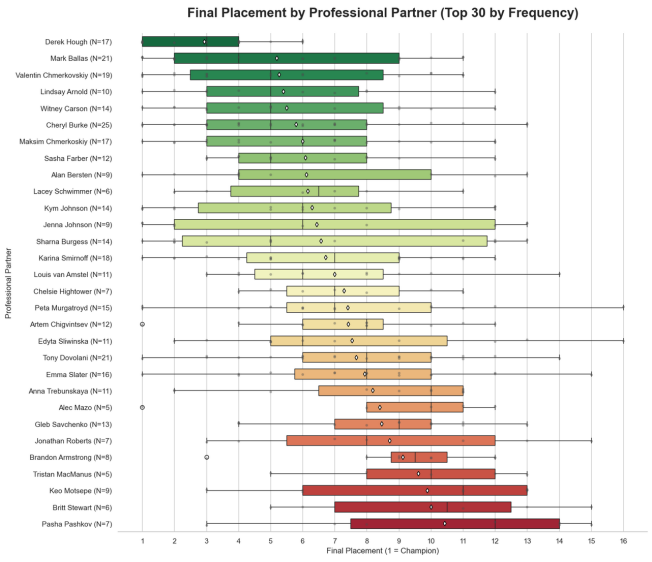

Figure 19: Partner Effect

(3) Core Industry Bonus

Industry background is a key determinant, accounting for 22.30% of total variance. Contestants from racing, social media, and athletics perform better on average.

(4) Age Decay Effect

Age negatively impacts rankings, with median positions shifting downward from <30 to >50, indicating that the probability of achieving high rankings declines significantly with increasing age.

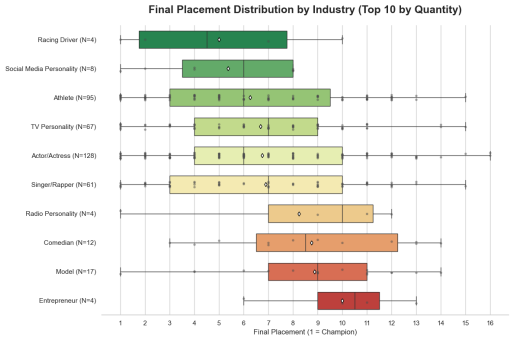

Figure 20: Industry Effect

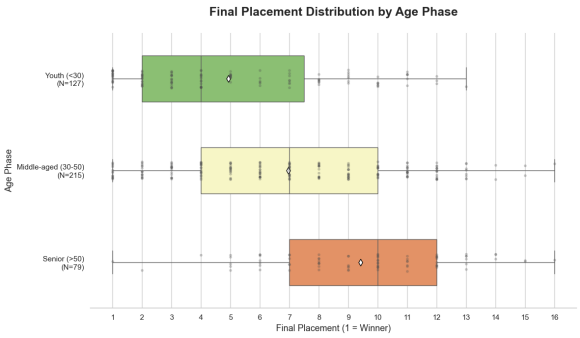

Figure 21: Age Effect

6.2.2 Path Preference Similarity and Divergence

(1) Dual Negative Consensus Along the Age Dimension

Age negatively affects both judges (-0.53) and fans (-0.35); judges are more sensitive, while fan responses are more variable, reflecting complex an esthetic preference dynamics across age stages.

(2) Synchronous Gain from Professional Partners

The partner effect shows strong, comparable positive impacts on both the judge (1.20) and fan (1.12) paths. Top professional partners, such as *Derek Hough* and *Mark Ballas*, enhance judges' scores through choreography and translate their personal appeal into fan votes, serving as one of the few stable factors bridging the evaluation gap.

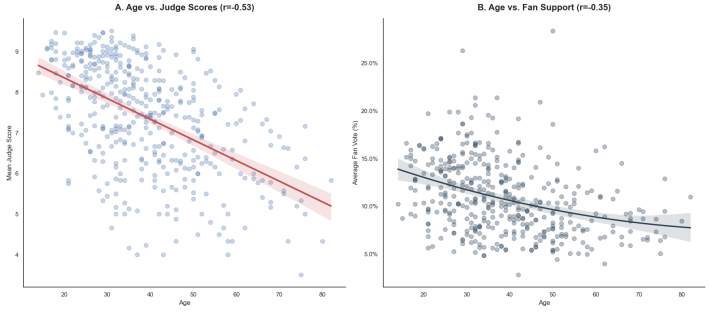

Figure 22: Similarity between Age

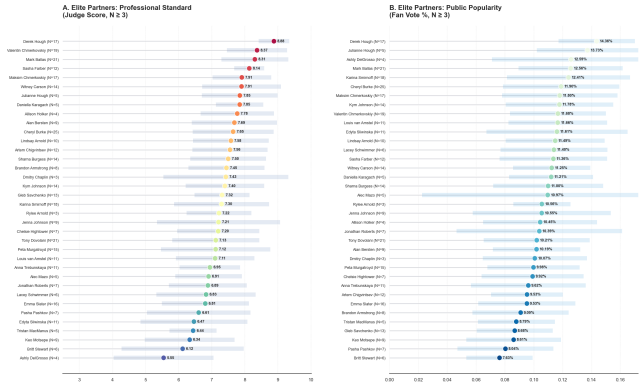

Figure 23: Similarity between Partners

(3) Industry Background Preference Difference

Industry background is a key positive predictor in both paths, but the fan path (2.02) assigns a higher premium than judge path (1.58), indicating that fan affiliation further amplifies the advantage conferred by certain professions, such as *athletes* and *singers*.

(4) Divergence in Evaluation Logic of Regional Characteristics

Regional attributes produce notable evaluation divergence: *Lat/Asi* contestants receive near-neutral judge scores (-0.03) but much lower fan votes (-0.56). Despite their technical proficiency, international contestants often encounter a distinct 'cultural alienation' during the domestic U.S. voting phase, leading to a significant decline in public popularity. A similar pattern is observed among contestants from the *EU* group.

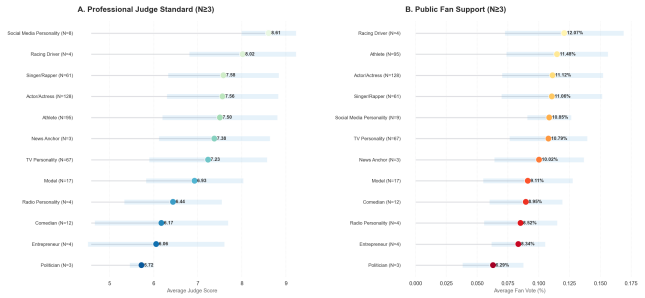

Figure 24: Divergence across Industries

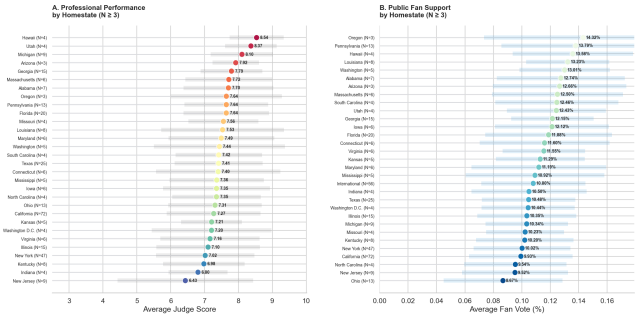

Figure 25: Divergence across Regions

(4) Synchronous Gain from Ballroom Partners

The partner effect shows strong, comparable positive impacts on both the judge (1.20) and fan (1.12) paths. Top professional partners, such as *Derek Hough* and *Mark Ballas*, enhance judges' scores through choreography and translate their personal appeal into fan votes, serving as one of the few stable factors bridging the evaluation gap.

6.2.3 Decision Power Allocation

The integrated regression layer abandons a single global average weighting scheme and instead performs segmented weight estimation across three historical competition phases, thereby precisely capturing the actual power allocation between judges and fans under different rule structures. The results are reported in the following table:

Table 4: Season range and corresponding mechanism weights

Season Range	Mechanism Used	Judge Weight	Fan Weight	Model R^2
1-2	Rank-based	67.79%	32.21%	0.3097
3-27	Percent-based	35.63%	64.37%	0.3506
28-34	Rank + Judge Save	52.93%	47.07%	0.3217

We find that, compared with the percent-based format, the Rank + Judge Save system significantly increases judges' influence and mitigates the bias toward fan voting observed in earlier

formats. By contrast, the pure rank-based format has a severely limited sample size, making its results lack general applicability.

7 Task 4: Log-Normal Z-Score Fusion System (LNZF System)

In response to the competition format biases identified earlier, this study proposes a score adjustment scheme based on statistical transformations, called **log-normal Z-score fusion (LNZF) system**.

7.1 Scoring Framework

The LNZS system applies the following three-step mathematical transformation, resolving the imbalance in evaluative power at the fundamental structural level:

(1) Judges' Scores Standardization

To address power dilution caused by extremely small variance in judges' scores, rescaling is applied to forcibly align evaluation scales, ensuring that even minor professional score differentials exert sufficient influence on decision-making.

$$Z_J = \frac{J - \mu_J}{\sigma_J} \tag{29}$$

(2) Fan Votes Logarithmic Standardization

To address extreme popularity effects arising from the power-law distribution of fan votes, nonlinear compression is applied to reduce the weight of extreme vote counts, thereby preventing the competition from degenerating into a purely popularity-driven contest.

$$Z_V = \frac{\log p - \mu_{\log p}}{\sigma_{\log p}} \tag{30}$$

(3) Dynamic Weight Evolution

We introduce a Sigmoid function to dynamically map weights based on the weekly dispersion of judges' scores, thereby smoothly balancing professionalism and entertainment and addressing the limitations of fixed weights in adapting to varying scoring characteristics.

$$\alpha(t) = 0.3 + \frac{0.5}{1 + e^{-k(\sigma_t - \sigma_{\text{median}})}} \tag{31}$$

Here, σ_t denotes the weekly score standard deviation, σ_{median} is the historical median, and sensitivity $k = 1$ ensures smooth transitions. A base value of 0.3 maintains judges' baseline authority, while a scaling factor of 0.5 caps their maximum weight at 0.8, preserving fan participation.

(4) Score Calculation

After transforming indicators and computing weights, the system calculates the final integrated score using a linear weighting formula.

$$S = \alpha(t) Z_J + (1 - \alpha(t)) Z_V \tag{32}$$

7.2 Monte Carlo Performance Analysis and Sensitivity Analysis

In order to evaluate the superiority of the Fusion model compared to the Rank and judge save combined method derived in Task 2, we conducted large-scale Monte Carlo simulations similar to Task 2.

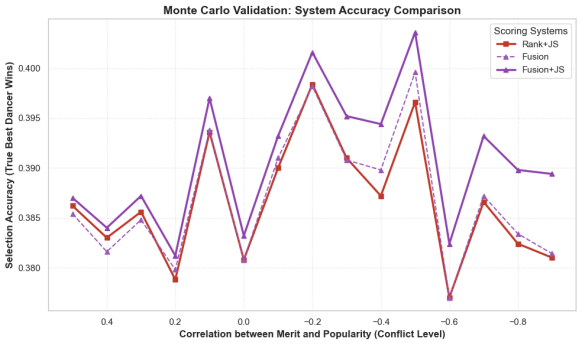

Figure 26: Selection Accuracy Comparison

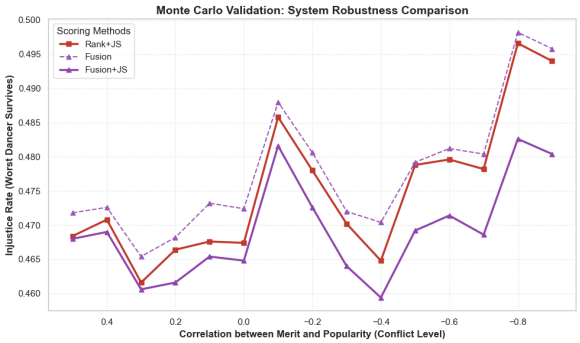

Figure 27: Injustice Rate Comparison

The results show that the LNZF system effectively eliminates contestants whose popularity far exceeds their actual performance while selecting truly professional contestants. By making fan influence bounded and controllable, it achieves a smoother transition between consensus weeks and controversy weeks, and aligns the final trophy distribution more closely with the contestants' technical abilities.

7.3 From Adaptive to Piecewise

Although the Sigmoid-based adaptive weighting mechanism is mathematically rigorous, this "black-box" weekly fluctuation is not conducive to fan understanding of the voting logic in actual TV show operations.

To balance mathematical accuracy with rule transparency, we propose a Piecewise Constant Weight System. In the first 5 weeks, when contestant performance levels are uneven and judges' professional selection is crucial, we set $\alpha = 0.6$. In the mid-stage (until just before the final), we set $\alpha = 0.5$. Finally, for the final, where we hope to select a champion more aligned with audience expectations, we set $\alpha = 0.4$.

It is worth noting that the experiment on outcome overlap shows that the consistency rate between the piecewise constant method and the complex adaptive Sigmoid model in determining elimination candidates reaches 95.8%.

8 Strengths and Weaknesses

8.1 Strengths

- **Integrated Analytical Pipeline:** The study constructs a coherent evaluation chain, transitioning from the inference of unobservable fan shares (MSFIM) to the diagnostic analysis of competition formats (Task 2), and concluding with the global factor attribution via MFIAM. This ensures a comprehensive capture of both initial social capital and performance dynamics.

- **Robust Latent Variable Inference:** By employing a Constrained Monte Carlo approach within a Bayesian framework, the model successfully back-calculates fan voting shares from historical outcomes. This effectively addresses the lack of public voting totals and provides a rigorous data foundation for subsequent bias analysis.
- **Dynamic and Adaptive Weighting:** The introduction of a Sigmoid-based weight evolution function allows judges' authority to adapt to weekly score dispersion. This mechanism achieves an optimal balance between technical rigor and audience engagement, mitigating the rigidity of fixed-weight scoring systems.
- **Multi-dimensional Attribution Logic:** The MFIAM leverages nested random effects to isolate "Partner Effects" and "Industry Genetics". This high-resolution profiling facilitates the precise quantification of a contestant's "Winning Potential" prior to the onset of the competition.

8.2 Weaknesses

- **Boundary Constraints of the Inference Model:** The MSFIM operates on the assumption that structural judge-fan discrepancies drive controversial outcomes. However, spontaneous shifts in voting behavior due to "viral moments" may occur within short intervals, which the model—restricted by historical data granularity—may smooth over as averaged trends.
- **Forced Aggregation in Geographic Analysis:** To maintain statistical robustness and avoid over-fitting, 25 subregions were aggregated into 4 macro-groups. This necessary simplification ensures the reliability of general conclusions regarding "Home-field Advantage" but may mask nuanced intra-regional cultural affinities.
- **Idealized Assumption of Format Bias:** Our diagnostic simulations assume consistent judicial scoring standards across seasons. In practice, judges may recalibrate their strictness based on the overall quality of a specific season's cast—a "production-driven bias" that is difficult to isolate without access to internal show production records.

References

[1] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.

[2] Metropolis, N., & Ulam, S. (1949). The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247), 335–341. http://www.jstor.org/stable/2280232

[3] Laird, N. M., & Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4), 963–974. https://doi.org/10.2307/2529876

[4] Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, 311(5762), 854–856. https://doi.org/10.1126/science.1121066

Memo

To: Dancing with the Stars (DWTS) Producers

From: MCM Team 2621213

Date: February 1, 2026

Subject: Recommendations for Scoring Mechanisms and Audience Engagement Strategies

Through a systematic analysis of 34 seasons, our research examined methods of combining judge and fan votes in different evaluation systems, identifying the system that fundamentally ensures the fairness of the competition. We found that the percent-based format systematically amplifies the influence of fan votes (quantified by our Fan Power Index, FPI), often leading to a "fan dictatorship" where contestants with strong professional abilities are surpassed by popularity-driven contestants. In contrast, the rank-based system provides a variance-stabilizing mechanism when FPI is high, maintaining the weight of professional judgment. Furthermore, our Multi-Factor Impact Assessment Model (MFIAM) indicates that contestant backgrounds (accounting for 36.13% of the variance) and the "brand equity" of professional partners (accounting for 12.59%) are significant pre-determinants of success.

Our simulations of controversial cases demonstrate that current systems without structural safeguards are vulnerable to popularity bias. The introduction of the *Judge Save mechanism* is the most effective intervention identified; by transitioning the final decision in elimination rounds to a qualitative expert judgment, the show filters out “popularity-only” advancements. Sensitivity analyses show that the “Rank + Judge Save” combination exhibits the highest robustness, maintaining an injustice rate below 40% even under extreme misalignment between judge and fan preferences, ensuring the trophy represents dance excellence.

To maximize both fairness and audience attraction, we propose the following strategies:

- **Strategic Casting and Cultural Balancing:** Use “Background Premium” data to balance the cast; pairing contestants facing an “Age Penalty” with elite pros can offset disadvantages and maintain demographic diversity. In addition, in order to enhance the attractiveness of the program, these "elite dance partners" should be utilized to drive those athletes who are facing a "cultural alienation" background (such as international athletes), thereby balancing the competitive landscape and expanding the global audience coverage of the program.
- **Adopting the "LNZF + Judge Save" system:** By using the Log-Normal Z-Score Fusion System (LNZF System) we proposed, it ensures that professional standards are protected from extreme popularity interference, which helps maintain the long-term credibility of the show.
- **Clarifying the selection goal:** Since this is a variety show rather than a professional competition, it is still uncertain whether the final champion must possess stronger professional skills. The evaluation system directly determines the characteristics of the final champion. Therefore, it is essential to first clarify the desired champion profile and then choose the most appropriate competition format based on this goal.

Report on Use of AI

1. Google *Gemini*(Dec 17, 2025 version, Gemini 3 Flash)
   Query1:<Explain the basic principles of the linear mixed effects model.>
   Output:<At its core, a **Linear Mixed Effects Model (LMM)** is a statistical framework used to analyze data that is "nested" or "clustered." While a standard linear regression assumes every data point is independent, LMMs recognize that some data points are naturally related—like multiple test scores from the same student or several leaf measurements from the same tree.

   Think of it as the "Goldilocks" approach: it doesn't treat everyone as exactly the same, but it doesn't treat everyone as a completely separate universe either.

   —

   1. The Two Components: Fixed vs. Random

   The "Mixed" in the name comes from the combination of two types of effects:

   * **Fixed Effects:** These are the factors you are primarily interested in. They represent the "population average." If you're testing a new drug, the dosage is a fixed effect. You expect it to have a consistent, predictable impact across the whole group. * **Random Effects:** These account for "noise" or variation that you can't control but need to acknowledge. They represent the differences between clusters (e.g., individual patients, specific cities, or different schools).

   —

   2. Why Use Them?

   Standard regression (Ordinary Least Squares) can fail when data is grouped because it violates the assumption of independence. LMMs solve several problems:

   * **Non-Independence:** It handles the fact that observations within a group are more similar to each other than to observations in other groups. * **Missing Data:** Unlike some older methods (like repeated measures ANOVA), LMMs are quite robust if some participants miss a few data points. * **Flexibility:** You can model both **Random Intercepts** (everyone starts at a different baseline) and **Random Slopes** (everyone reacts to the treatment at a different rate).

   —

   3. The Mathematical Intuition

   In a standard regression, we use . In a Mixed Model, we add a term for the random cluster-specific deviation ():

   Where:

   * : The **Fixed Effects** (the overall trend). * : The **Random Effects** (how much a specific group deviates from that trend). * : The **Residual Error** (the remaining unexplained "jitter").

   —

   4. Random Intercepts vs. Random Slopes

To visualize this, imagine tracking the growth of different puppies:

| Feature | Description | | — | — | | **Random Intercept** | Each puppy starts at a different weight (some are Chihuahuas, some are Great Danes), but they all grow at the same average rate. | | **Random Slope** | Each puppy grows at a different speed (some sprout quickly, others slowly) regardless of their starting weight. |

—

Summary

Linear Mixed Effects Models allow you to zoom out to see the big picture (Fixed Effects) while acknowledging the unique "quirks" of your individual groups (Random Effects). They are the industry standard for longitudinal studies, ecology, and social sciences.

Would you like me to walk through a specific example in R or Python to show you how to code one of these models?>

2. Google *Gemini*(Dec 17, 2025 version, Gemini 3 Flash)
Query2:<Simplify the following content: It is noteworthy that the paradoxical widening of confidence intervals (CI) in the final stages stems from increased statistical sensitivity due to sample size compression and score compression. As the number of contestants decreases, each percentage point carries more weight. Simultaneously, as elite contestants' judges' scores converge, the model must accommodate a broader range of fan vote shares to justify all logically possible rankings. While the absolute CI expands, the Coefficient of Variation (CV) typically continues to decline, proving that the model's relative inference precision is actually optimizing as it scales with the higher mean vote shares.>
Output:<To put it simply, the data gets "noisier" on the surface while actually becoming more accurate underneath. Here is the breakdown of why those confidence intervals (CIs) are acting so strangely in the finale:

1. The "Big Fish, Small Pond" Effect

As contestants are eliminated, the sample size shrinks. In a smaller group, every single percentage point matters more. If you lose just a few fans, it swings your standing much more violently than it did in the early rounds.

2. The Scoring Bottleneck

By the finale, the judges are usually giving everyone "elite" scores (e.g., 9s and 10s). Because the judges' scores are so close together, the model has to look at a much wider range of potential fan votes to figure out who could actually win. This "uncertainty" makes the error bars (CI) look wider.

3. Precision vs. Scale

Even though the range of the prediction looks wider, the **Coefficient of Variation (CV)**—which measures the "error" relative to the "total size"—is actually dropping.

Think of it like this:

* **Early Rounds:** You are guessing if someone gets 5% or 7% of the vote (a 2% gap). * **Finale:** You are guessing if they get 40% or 45% (a 5% gap).

The 5% gap looks "bigger" (wider CI), but because the total number (40%) is so much higher, the prediction is mathematically more stable and precise.

—

Summary Table

| Metric | What's happening? | Why? | | — | — | — | | **Confidence Interval** | Widens (Looks worse) | Fewer people and tie-breaking judge scores. | | **Coefficient of Variation** | Shrinks (Is better) | The model is more confident relative to the high vote volume. |

—

Would you like me to rewrite this for a specific audience, such as a casual fan blog or a technical research paper?>

3. Google *Gemini*(Dec 17, 2025 version, Gemini 3 Flash)
Query3:<Introduce the dance program Dancing with the Stars (DWTS).>
Output:<If you're looking for the ultimate blend of sequins, spray tans, and high-stakes footwork, you've found it. **Dancing with the Stars (DWTS)** is the long-running reality competition that pairs professional ballroom dancers with celebrities to see who can master the art of the Argentinian Tango without tripping over their own feet.

—

The Core Concept

The show is based on the British hit *Strictly Come Dancing*. The premise is simple but addictive:

* **The Pairings:** An A-to-D-list celebrity (athletes, actors, reality stars, or even politicians) is paired with a world-class professional ballroom dancer. * **The Grind:** Each week, the couples learn a new style—ranging from the elegant **Waltz** to the high-octane **Jive**. * **The Judgment:** Performances are scored by a panel of expert judges and then combined with viewer votes to determine who stays and who is "eliminated" from the ballroom.

Why People Watch

It's more than just dancing; it's a masterclass in "the journey." Fans tune in for:

1. **The Transformation:** Seeing an NFL linebacker who claims to have "two left feet" suddenly execute a graceful Foxtrot is genuinely impressive. 2. **The Glamour:** The costume department is legendary. We're talking thousands of crystals per dress and more spandex than a 1980s aerobics video. 3. **The Drama:** Between the grueling rehearsal schedules, occasional injuries, and the "will-they-won't-they" chemistry between partners, there's never a dull moment.

The Trophy

The ultimate prize isn't a million-dollar contract; it's the **Mirrorball Trophy**. While it might look like a sparkly prop from a 1970s disco, in the world of DWTS, it is the highest honor.

| Feature | Description | | — | — | | **Network** | ABC (and Disney+) | | **Judging Criteria** | Technique, Performance, and Chemistry | | **Fan Involvement** | Live voting heavily influences the results |

—

Would you like me to look up who the current reigning champion is or give you a list of the most iconic winners from past seasons?>

4. Google *Gemini*(Dec 17, 2025 version, Gemini 3 Flash)
Query4:<Formula for calculating coefficient of variation.>
Output:<The **Coefficient of Variation (CV)** is a statistical measure used to determine the relative dispersion of data points in a data series around the mean.

It is particularly useful when you want to compare the degree of variation between two data sets that have different units or widely different means. Essentially, it tells you how large the standard deviation is relative to the average.

The Formula

To calculate the CV, you divide the standard deviation by the mean and typically multiply by 100 to express it as a percentage.

Where:

* **** (sigma) = Standard Deviation * **** (mu) = Mean

—

Why use CV instead of Standard Deviation?

While standard deviation gives us a sense of "spread," it is an absolute measure. If you are comparing the price volatility of a stock that costs $10 versus one that costs $1,000, a $5 standard deviation means something very different for each.

| Feature | Standard Deviation | Coefficient of Variation | | — | — | — | | **Type of Measure** | Absolute | Relative (Percentage) | | **Units** | Same as the data (e.g., kilograms, dollars) | Unitless | | **Best Use Case** | Understanding spread within one dataset | Comparing consistency across different datasets |

> **Note:** The CV should only be used for data measured on a **ratio scale** (data with a true zero, like height, weight, or Kelvin temperature). Using it on interval data (like Celsius or Fahrenheit) can lead to misleading results because the "zero" point is arbitrary.

Would you like me to walk through a step-by-step example calculation using a specific set of numbers?>