

1 评分数据的“堆积”与方差压缩—实证分析

1.1 分析目的与数据说明

本节旨在对评审评分分布进行实证分析，检验是否存在以下现象：

1. **得分膨胀 (Score Inflation)**: 随赛程推进，评分向高端量表（7–10 分）偏移。
2. **方差压缩 (Variance Compression)**: 随赛程推进，评分的有效方差收缩，即“天花板效应 (Ceiling Effect)”。

数据来源: 使用 2026_MCM_Problem_C_Data.csv 数据集。提取评委单分列 `weekX_judgeY_score` (1–10 分)，仅保留 `score > 0` 且非 N/A 的有效评委分 (0 表示该选手已淘汰)。数据按 **周序 (Week 1–11)** 汇总，跨所有赛季合并 (同一周序在不同赛季均视为同一“周”)。最终得到 **9,112** 条有效评委分样本。

1.2 汇总表：按周的均值、标准差、方差与变异系数

表 1 展示了每周的有效评委分数个数、均值、标准差、方差和变异系数 ($CV = \sigma/\mu \times 100\%$)。

表 1: 按周序汇总的评分统计量

Week	Count (n)	Mean (μ)	Std (σ)	Var (σ^2)	CV (%)
1	1330	6.49	1.27	1.62	19.58
2	1299	6.94	1.14	1.29	16.35
3	1187	7.39	1.23	1.52	16.65
4	1081	7.81	1.21	1.47	15.52
5	951	8.21	1.17	1.37	14.26
6	882	8.72	1.27	1.61	14.56
7	768	8.84	0.98	0.96	11.08
8	628	9.08	1.01	1.01	11.07
9	512	9.33	0.84	0.70	8.97
10	355	9.49	0.59	0.34	6.18
11	119	9.72	0.39	0.16	4.05

对表 1 的解读：

- **得分膨胀**：周均分由第 1 周的 **6.49** 分单调上升至第 11 周的 **9.72** 分，整体向 7–10 分区间集中，与“随赛程推进评分向高端偏移”的描述一致。

- **方差压缩**: 标准差由第 1 周的 **1.27** 降至第 11 周的 **0.39**; 方差由 **1.62** 降至 **0.16**; 变异系数由 **19.58%** 降至 **4.05%**。后期周内评分高度趋同, 有效方差明显收缩, 存在明显的天花板效应。
- **样本量说明**: 早期周 (1-3) 约 1,200-1,300 条/周, 后期周 (10-11) 因赛程缩短与淘汰, 样本量减少 (355、119), 但趋势与中段周一致, 结论具有鲁棒性。

1.3 显著性检验：趋势是否具有统计意义

为了验证趋势的统计显著性, 我们对 **周序 (week)** 与 **周均分 (mean)**、**周内标准差 (std)** 进行了相关性检验与回归分析, 结果如表 2 所示。

表 2: 得分膨胀与方差压缩的统计显著性检验结果

检验项目	统计结果	解读
均分 ~ 周序 (Spearman)	$r = 1.00, p \approx 0.000$	周序与均分完全单调正相关, 得分膨胀显著 。
均分 ~ 周序 (OLS 斜率)	$slope = 0.32 \text{ 分/周}, p \approx 0.000$	平均每周约上升 0.32 分, 趋势显著。
标准差 ~ 周序 (Spearman)	$r = -0.85, p = 0.001$	周序与标准差显著负相关, 方差压缩显著 。
标准差 ~ 周序 (OLS 斜率)	$slope = -0.08 \text{ 分/周}, p = 0.0006$	周内离散度随周序显著下降。

结论: 从量化角度分析, 数据中存在显著的 **** 得分膨胀 **** 与 **** 方差压缩 **** 现象 ($p < 0.01$)。后期周评分集中在高分区间, σ^2 明显变小。这证实了“当裁判评分高度趋同导致 $\sigma_j^2 \rightarrow 0$ 时, 最终判别力将更多由方差更大的分量 (如观众投票) 主导”的论断。

1.4 图表与解读

以下图表进一步直观展示了上述分析结果。

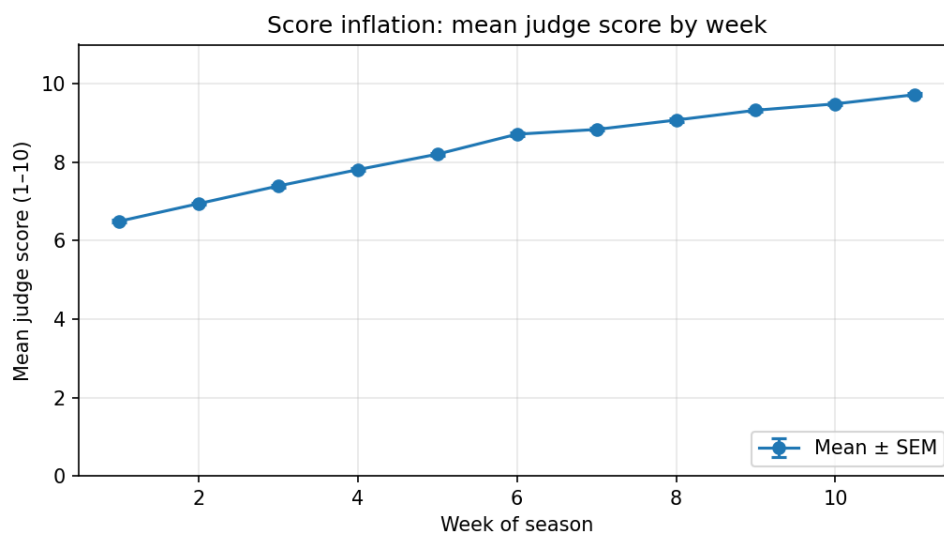


图 1: **周均分随周序变化 (得分膨胀)**。横轴为周序, 纵轴为均值。误差棒表示标准误 (SEM)。均分呈现单调上升趋势。

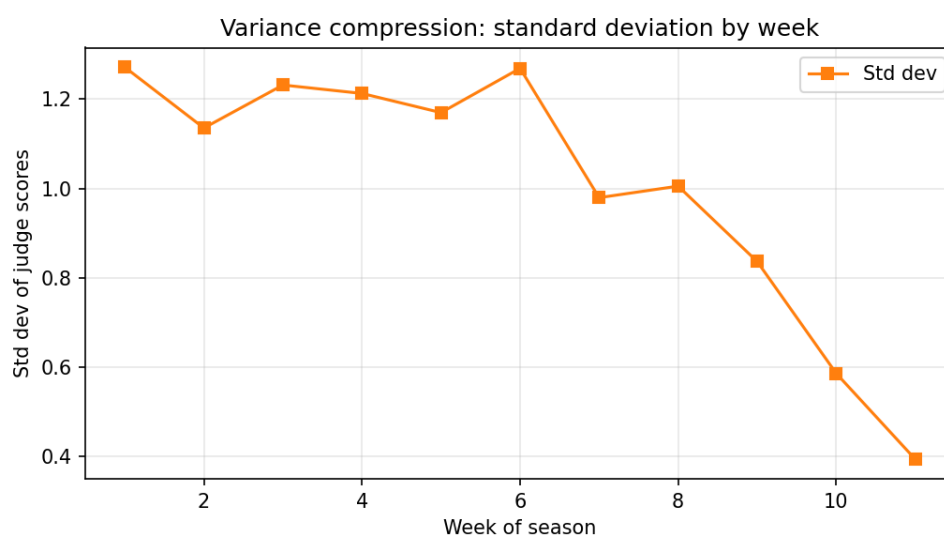


图 2: **周内标准差随周序变化 (方差压缩)**。横轴为周序, 纵轴为标准差。随着比赛进行, 评分离散度显著下降。

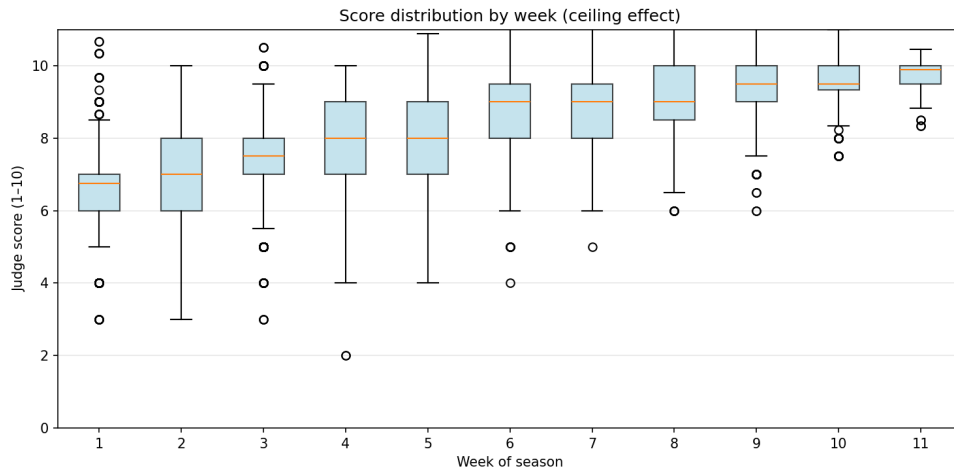


图 3: 每周得分分布箱线图。随周序增加, 箱体整体上移且变窄, 高分端堆积明显, 低分端几乎消失, 体现了“天花板效应”。

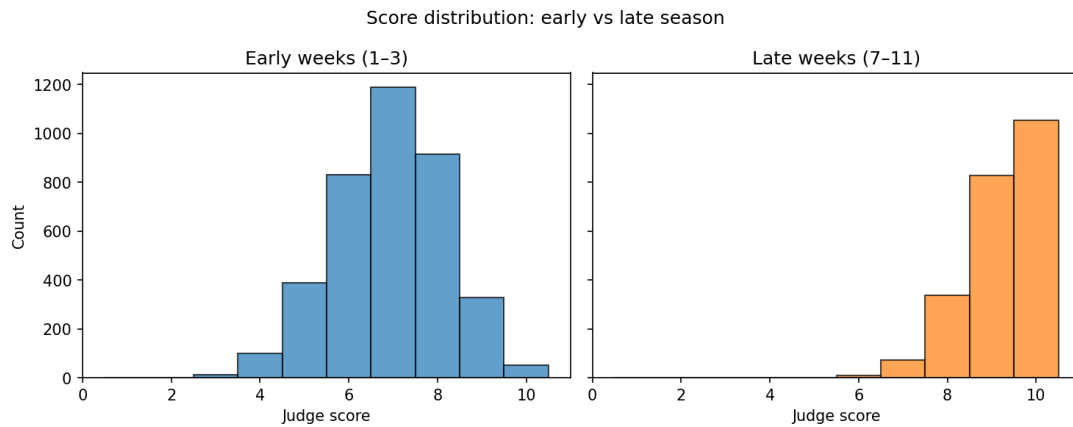


图 4: 早期周 (1-3) vs 后期周 (7-11) 得分直方图对比。左图分散, 右图集中于 8-10 分区间。

1.5 小结与对文中论断的对应

本实证分析结果与模型假设及理论推导高度吻合, 对应关系总结如下 (见表 3):

表 3: 文中论断与实证分析的对应关系

文中表述	本实证分析对应结果
得分膨胀; 评分向 7-10 分偏移	表 1 显示均分 $6.49 \rightarrow 9.72$; OLS 回归斜率 ≈ 0.32 分/周 ($p \approx 0$)。
方差压缩; 天花板效应	标准差 $1.27 \rightarrow 0.39$, 方差 $1.62 \rightarrow 0.16$; Spearman 相关系数 $r = -0.85$ ($p = 0.001$)。
评分趋同导致 $\sigma_j^2 \rightarrow 0$, 判别力下降	后期周 CV 降至约 4-6%, 方差显著缩小, 验证了评委评分在总分中有效权重的下降趋势。

最终结论: 本数据集呈现出具有统计显著性的 ** 得分膨胀 ** 与 ** 方差压缩 ** 特征。表 1 与图 1-4 为文中关于“堆积”与“权重失效”的讨论提供了坚实的实证基础。