

山东省 2029 年新高考五级赋分制 测验等值与分数分布预测报告

Xinyu Yang

Email: Pana.Yang@hotmail.com | ORCID: 0009-0007-2600-0948 | DOI: 10.5281/zenodo.22226943

完成日期: 2026 年 9 月 1 日 | 代码与数据仓库: rad:zYrJE5TYBeoGbNj2Ny4quMLGauMw

摘要: 根据教育部关于深化新高考等级赋分制改革的统一部署, 山东省普通高校招生夏季高考自 2029 年起正式由现行“8 等级赋分制”(A、B+、B、C+、C、D+、D、E 共 8 级, 区间 21–100 分) 并轨转换为国家统一部署的“5 等级赋分制”(A、B、C、D、E 共 5 级, 区间 30–100 分)。本文基于山东省 2020–2026 年连续 7 年官方一分一段全样本实证数据(涵盖 438 万人次大样本)与《山东统计年鉴》2002–2016 年学龄人口出生队列数据, 构建了结合教育考试测量学(Psychometrics)与微观计量经济学的测验等值(Equating)与分数分布预测模型体系。

本研究系统实现了 16 项核心数理模型与控制机制: (1) 学龄人口队列递推与教育转化漏斗模型(测定 2029 届参考规模为 74.82 万人); (2) 现行 8 等级制向 2029 新 5 等级制并轨的非线性量表变形映射; (3) 2024 年高校专业选考“物化强绑定”政策断点的多层级分层计量识别; (4) 高分段前 50 名脱敏排位的局域对数极值 Pareto 逆回归与 C^1 平滑连接; (5) 局域多项式近邻选考人次全域恒等守恒律($\sum_{s=1}^6 P_s \equiv 3.0$); (6) Fritsch-Carlson 单调保序三次 Hermite 插值多项式(PCHIP)连续化; (7) 7 维学科潜能向量多元正态相依结构标定($\theta \sim \mathcal{N}_7(\mathbf{0}, \mathbf{R})$); (8) 选考单科赋分期望中立性定理($\mathbb{E}[T_s] \equiv 68.32$ 分); (9) 20 种选科组合信息论最小交叉熵规划(I-Divergence / SLSQP); (10) 考生自选择 Multinomial Logit 与带阻尼($\eta = 0.70$) IPF 影子价格均衡系统; (11) 统考语数外潜能全域法定支持域 $[0, 450]$ 广义 Logistic 映射与微观破并列微扰($\sigma_{\text{jitter}} = 0.22$ 分); (12) 非参数面板分位数弹性回归 $\hat{\beta}(p)$ 与规模对数正交修正; (13) 历年命题难度冲击的分位数等百分位单调保序包络投影($\Pi_{\leq 0}, \Pi_{\geq 0}$); (14) 法定五级阶梯赋分公式与单科独立取整方差守恒($\text{Var}(e_{\text{elec}}) = 0.25$); (15) 注入宏观考试命题与选科漂移方差($\sigma_{\text{macro}} = 3.5$ 分)的保序回归(PAVA) 95% 全域联立置信走廊; (16) 基于全方差定理的 ANOVA 预测总不确定性正交分解($\sigma_{\text{total}}^2 = 20.98 \text{分}^2$)。

预测表明, 在 2029 年基准常态命题情景下, 全省总分均值上升至 481.05 ± 0.10 分(较历史 7 年池化均值净上移 +47.76 分), 理论等值排位基准线为: 特控线(Top 20%) 558.00 分、一段线(Top 44%) 498.00 分、二段线(Top 95%) 332.00 分, 四分位距 IQR 为 122.00 分。

特别说明与免责声明: 本报告所测算之一段线、特控线、二段线等系依据历年常态位次百分位(Percentile Cutoffs)反推之测验等值理论基准线。实际招生录取中的各批次控制线与高校投档线严格取决于教育部当年下达之法定招生计划数、高校专业选拔策略及招生管理部门政策, 本报告测算结果仅供教育测量学术研究与等值分析参考。

关键词: 高考改革; 等级赋分制; 测验等值; 人口队列递推; 物化绑定; 微观计量模型; 保序回归; 方差分解

目录

1 问题定义与政策背景	2
1.1 现行八等级制与 2029 新五等级制变革对照	2
1.2 2029 届考生规模的人口统计学队列递推与教育转化漏斗模型	2
1.3 专业选考“物化强绑定”政策断点(2024 突变)与多层级计量识别策略	3
1.4 测验等值(Psychometric Equating)的核心计量难点	4
1.5 测验学数学记号与符号约定	4
2 历史数据治理与微观实证诊断(2020–2026)	4
2.1 高分段脱敏 Pareto 尾部外推与 PCHIP 连续化方程	4
2.2 历史全量数据实证诊断大表(2020–2026)	4
3 微观计量经济学与高维潜能相依模型	5
3.1 7 维潜能向量与截断多元正态矩匹配反解标定	5
3.2 选科组合信息论最小交叉熵规划(I-Divergence KL-Minimization)	5
3.3 选科自选择排序效应与带阻尼 IPF 迭代校正	6
4 全域法定统考映射与法定五级阶梯赋分	6
4.1 全域法定支持域统考潜能映射与测验微扰	6
4.2 卷面命题难度冲击的分位数等百分位单调保序包络投影	7
4.3 法定五级赋分阶梯公式与离散格点量化方差守恒	7
5 三大命题情景、置信走廊与误差预算	8

5.1 保序回归 (PAVA) 95% 全域联立置信走廊

5.2 预测总方差不确定性预算与 ANOVA 正交分解

5.3 历年位次标准化演进与总方差分解全景

5.4 2029 核心控制线与测验统计量权威汇总大表

6 结论与志愿填报战略建议

7 参考文献 (References)

8

8

9

9

10

10

1 问题定义与政策背景

1.1 现行八等级制与 2029 新五等级制变革对照

自 2020 年山东新高考综合改革落地起，统考科目（语数外各 150 分，满分 450 分）计原始卷面分；6 门等级考选考科目（物理、化学、生物、政治、历史、地理）执行**现行 8 等级赋分制**，沿用至 2028 年。自 2029 年夏季高考起，等级考科目正式并轨精简为**5 等级赋分制**。

1. 现行 8 等级赋分标准（2020–2028 年）：
- 分段与比例：按正态分布原则将考生原始分由高到低划分为 8 个等级：A (3%)、B+ (7%)、B (16%)、C+ (24%)、C (24%)、D+ (16%)、D (7%)、E (3%)；

· 赋分区间：依次对应转换为 91–100 分、81–90 分、71–80 分、61–70 分、51–60 分、41–50 分、31–40 分、21–30 分，各等级均为 10 分等宽，最低起点为 21 分；

· 转换公式：设某等级区间原始分为 $[Y_2, Y_1]$ ，赋分区间为 $[T_2, T_1]$ ，原始分 y 对应的等级分 t 满足：

$$\frac{Y_1 - y}{Y_1 - Y_2} = \frac{T_1 - t}{T_1 - T_2}$$

2. 2029 年及以后新 5 等级赋分标准（新调整）：
- 分段与比例：精简为 5 个等级：A (15%)、B (35%)、C (35%)、D (13%)、E (2%)；

· 赋分区间：依次对应转换为 100–86 分（A 等, 跨度 14 分）、85–71 分（B 等, 跨度 14 分）、70–56 分（C 等, 跨度 14 分）、55–41 分（D 等, 跨度 14 分）、40–30 分（E 等, 跨度 10 分，最低保底 30 分）。

新五级	新占比	新 赋 分 区 间	新跨度	现行八级对应	现行占比	现行区间	新旧对应结构解析
A 等	15%	100 – 86 分	14 分	A(3%) + B+(7%) + B(前 5%)	15%	100 – 76 分	A 等容积扩大 5 倍，高分段量表弹性释放
B 等	35%	85 – 71 分	14 分	B(后 11%) + C+(24%)	35%	75 – 61 分	中上段赋分基准普遍上移 10 分
C 等	35%	70 – 56 分	14 分	C(24%) + D+(前 11%)	35%	60 – 46 分	中段赋分基准上移 10–15 分
D 等	13%	55 – 41 分	14 分	D+(后 5%) + D(7%) + E(前 1%)	13%	45 – 28 分	中低段赋分基准大幅抬升
E 等	2%	40 – 30 分	10 分	E(后 2%)	2%	27 – 21 分	底限由 21 分提升至 30 分保底

1.2 2029 届考生规模的人口统计学队列递推与教育转化漏斗模型

2029 年夏季高考考生的常规基准出生队列为 **2011 年** 出生人口。根据《山东统计年鉴》官方人口统计序列（2002 年 101 万、2003 年 104 万、2004 年 114 万、2005 年 112 万、2006 年 108 万、2007 年 104 万、2008 年 106 万、2009 年 110 万、2010 年 111 万、**2011 年 110.0 万人**，至 2016 年二孩历史峰值 177 万人），建立学龄人口队列递推与教育转化漏斗方程（Cohort Component Progression Model）：

$$N_{2029} = B_{2011} \times s_{\text{surv}} \times s_{\text{comp}} \times r_{\text{highschool}} \times r_{\text{gaokao}} \times (1 + r_{\text{repeat}})$$

式中各结构性转移概率界定如下：

1. 出生队列基数： $B_{2011} = 1,100,000$ 人；

2. 0-18 岁累计存活率：依据山东省人口生命表生命期望外推测定 $s_{\text{surv}} = 0.993$ ；
3. 九年义务教育巩固率：依据《山东省教育事业统计公报》测定 $s_{\text{comp}} = 0.985$ ；
4. 初中升入普通高中录取率（普高分流比）：结合山东省普通高中扩招实绩（普通高中在校生由 2023 年 198.54 万人增至 2025 年 215.80 万人），测定 $r_{\text{highschool}} = 0.63357$ ；
5. 高中应届生夏季高考有效参考率：测定 $r_{\text{gaokao}} = 0.980$ ；
6. 复读生（往届生）稳态回流增量比：测定稳态比率 $r_{\text{repeat}} = 0.120$ 。

代入参数求得 2029 年夏季高考全省有效总参考人数为：

$$N_{2029} = 1,100,000 \times 0.993 \times 0.985 \times 0.63357 \times 0.980 \times (1 + 0.120) = 748,200 \text{ 人} \quad (74.82 \text{ 万人})$$



图 1 图 1：山东省高考（2020–2034 年）学龄人口队列递推、教育转化漏斗与生源规模长期预测全景图

1.3 专业选考“物化强绑定”政策断点（2024 突变）与多层级计量识别策略

山东省高考历史实证数据经历了一次重大的外生政策结构突变（Structural Break）：

1. 2020–2023 届（旧版选考指引期）：高校绝大多数理工农医专业仅要求“物理或化学选考一门即可”或不设限。考生中普遍存在“避难就易”博弈（大量出现“选物理弃化学”的物生地、物生政组合），导致化学选考人数偏低（全省仅 30%–35%），化学选考生源被高度提纯，物理与化学重叠度显著偏低。然而，该时期考生选科更纯粹地反映了个体学科兴趣与比较优势，为学科内内在难度基准与文理潜能原始相依结构的标定提供了无政策扭曲的干净先验信息；
2. 2024 届起（新版选考指引强绑定落地）：教育部明确规定 90% 以上理工农医专业必须“物理和化学同时选考”。全省选科格局剧烈洗牌：物理与化学选考率双双陡增至 50%–60%，“物化同选组合”（物化生、物化地、物化政）占全体考生比例跃升至 45% 以上，高学术潜能生源在理科高度极化集聚。

2020–2023 与 2024–2026 数据的分层计量经济学识别与协同利用策略：我们建立了结构性分层识别框架（Stratified Econometric Identification），精准发挥不同历史周期的信息优势：

- 学科内测验等值与潜能协方差标定（全量 7 年面板）：在第 3 章 7 维潜能相关矩阵 R 的矩匹配估计中，充分利用 2020–2026 全周期 7 年数据，借助 2024 年政策断点前后的自然实验外生变异（Natural Experiment Variation），将考生固有学科能力相关性与政策引致的选科博弈效应精准正交剥离；
- 生源规模对数弹性回归 $\hat{\beta}(p)$ （全量 7 年面板）：生源规模从 2020 年 52.5 万扩大至 2026 年 70.9 万的大跨度样本方差，为估计非参数分位数弹性提供了不可或缺的长序列自由度；
- 极端命题难度冲击包络边界（2020 最易卷与 2022 最难卷）：第 4.2 节平易情景以 2020 年全省高分均分卷面为上界，高难情景以 2022 年全省公认高难卷面（数学等科目高认知负荷）为下界，充分继承了 2020–2023 年实测极端冲击包络；

- 宏观选科组合外推与常态命题基准（后捆绑子期截断）：仅在 2029 届 20 种选科组合目标选考率 P_{target} 的外推及基准池化曲线 $Q^*(p)$ 处，严格截取 2024–2026 新规周期，防止将旧规下化学低选考率的制度性偏误线性外推至 2029 年。

1.4 测验等值（Psychometric Equating）的核心计量难点

- 等第跨度非等宽扭曲：现行制各级均为 10 分等宽，新制 A/B/C/D 为 14 分宽，E 等为 10 分宽，导致原始量表发生非线性拉伸；
- 选科生源自选择偏倚（Sorting Bias）：2024 年新规后高潜能考生向理科极化集聚；
- 生源规模对数稀释效应（Scale Elasticity）：考生总规模由 2020 年的 52.5 万人增至 2029 年的 74.8 万人，导致中低分段密度非线性稀释；
- 离散格点取整方差守恒：单科四舍五入离散化（ $\text{Var}(e) = 0.25$ ）与分数连续分布的数学兼容性。

1.5 测验学数学记号与符号约定

- 记 N 为全省考生总规模， $S \in [150, 750]$ 为高考文化总分；
- 记 $p \in [0, 1]$ 为上侧累积超越概率（Exceedance Fraction / CCDF），满足 $p = \frac{C(S)}{N}$ ，其中 $C(S) = \sum_{x=S}^{750} \text{count}(x)$ 为总分大于等于 S 的全省累积排位人数；
- 记 $Q(p) \equiv \bar{F}^{-1}(p) = F^{-1}(1-p)$ 为对应上侧分位数 p 的等值分数， $Q(0) = 750, Q(1) = 150$ 。

2 历史数据治理与微观实证诊断（2020–2026）

2.1 高分段脱敏 Pareto 尾部外推与 PCHIP 连续化方程

山东招考院对各科前 $n_{\text{masked}} = 50$ 名考生的分数与科目进行脱敏屏蔽。为避免全局插补引起的文科上翘偏倚与选考人次丢失，我们在近邻局域窗口 $[660, 680]$ 分对未脱敏排位 $r \in [n_{\text{masked}}, 200]$ 拟合对数极值 Pareto 逆回归：

$$S(r) = \alpha - \beta_{\text{tail}} \ln(r) + \varepsilon_r$$

外推尾部排位 $r \in [1, n_{\text{masked}}]$ 的等值连续分数为：

$$\hat{S}_{\text{tail}}(r) = \min(750.0, S_{\text{top}} - \hat{\beta}_{\text{tail}}(\ln(r) - \ln(n_{\text{masked}})))$$

在连接点 $r = n_{\text{masked}} = 50$ 处满足一阶平滑条件： $\hat{\beta}_{\text{tail}} = -50 \cdot \left(\frac{dQ(p)}{dr} \right) \Big|_{r=50^+}$ 。

在离散分位数网格 $p_k \in (0, 1)$ （取 $M = 20,000$ 个等距节点）上，采用 Fritsch-Carlson 单调保序三次 Hermite 插值多项式（PCHIP）构造全域连续分位数函数 $Q(p)$ ：

$$Q(p) = h_{00}(t)y_k + h_{10}(t)\Delta_k d_k + h_{01}(t)y_{k+1} + h_{11}(t)\Delta_k d_{k+1}, \quad t = \frac{p - p_k}{p_{k+1} - p_k}$$

其中三次 Hermite 基函数显式定义为：

$$h_{00}(t) = 2t^3 - 3t^2 + 1, \quad h_{10}(t) = t^3 - 2t^2 + t, \quad h_{01}(t) = -2t^3 + 3t^2, \quad h_{11}(t) = t^3 - t^2, \quad t \in [0, 1]$$

其一阶导数 d_k 严格满足 Fritsch-Carlson (1980) 对称调和平均单调性约束：

$$d_k = \begin{cases} \frac{2\Delta_{k-1}\Delta_k}{\Delta_{k-1} + \Delta_k} & \text{若 } \Delta_{k-1}\Delta_k > 0 \\ 0 & \text{若 } \Delta_{k-1}\Delta_k \leq 0 \end{cases}$$

对脱敏区间的 6 门选考科目人数，施加严格的选考人次总和守恒律：

$$\sum_{s=1}^6 P_{s(S)} \equiv 3.000, \quad \forall S \in [680, 750]$$

2.2 历史全量数据实证诊断大表（2020–2026）

年份	考生规模 N	均值 $\mu \pm \text{SE}$	标准差 σ	中位 P_{50}	四分位距 IQR	特控线 (Top20%)	一段线 (Top44%)	二段线 (Top95%)
----	-------------	------------------------	--------------	-------------	-------------	-----------------	-----------------	-----------------

2020	524,825	445.35 ± 0.14	100.53	454.0	148.0	535.0 分	470.0 分	273.0 分
2021	544,424	435.06 ± 0.13	96.83	446.0	142.0	520.0 分	461.0 分	268.0 分
2022	594,771	431.41 ± 0.12	93.83	438.0	137.0	514.0 分	453.0 分	272.0 分
2023	662,199	428.43 ± 0.12	99.33	433.0	145.0	516.0 分	450.0 分	262.0 分
2024	666,000	430.19 ± 0.12	99.91	437.0	149.0	520.0 分	454.0 分	261.0 分
2025	681,127	429.69 ± 0.12	102.76	438.0	152.0	522.0 分	455.0 分	252.0 分
2026	708,817	435.28 ± 0.12	101.68	442.0	150.0	527.0 分	459.0 分	262.0 分
7 年池化	626,023	433.29 ± 0.05	99.59	441.0	147.0	522.0 分	457.0 分	263.0 分

3 微观计量经济学与高维潜能相依模型

3.1 7 维潜能向量与截断多元正态矩匹配反解标定

设考生 $i \in \{1, \dots, N\}$ 的潜能向量服从 7 维多元正态分布：

$$\boldsymbol{\theta}_i = (\theta_{i, \text{comp}}, \theta_{i,1}, \dots, \theta_{i,6})^T \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$$

协方差矩阵 $\mathbf{R} \in \mathbb{R}^{7 \times 7}$ 反映不同学科间的潜在能力相关性，由历史选科生源均分偏倚矩匹配反解标定：

$$\min_{\mathbf{R} \succ 0, R_{jj}=1} \sum_{s=1}^6 (\Delta \bar{S}_s^{\text{obs}} - \Delta \bar{S}_s^{\text{sim}}(\mathbf{R}))^2, \quad \Delta \bar{S}_s = \sum_{m \in \mathcal{C}_s} w_m \mathbb{E} \left[\theta_{i,s} \mid U_{i,m} = \max_j U_{i,j} \right]$$

标定得到正定对称相关矩阵并实施 Cholesky 分解 $\mathbf{R} = \mathbf{L}\mathbf{L}^T$ ：

$$\mathbf{R} = \begin{pmatrix} 1.000 & 0.732 & 0.705 & 0.591 & 0.472 & 0.448 & 0.516 \\ 0.732 & 1.000 & 0.784 & 0.635 & 0.362 & 0.331 & 0.428 \\ 0.705 & 0.784 & 1.000 & 0.662 & 0.375 & 0.346 & 0.449 \\ 0.591 & 0.635 & 0.662 & 1.000 & 0.458 & 0.421 & 0.505 \\ 0.472 & 0.362 & 0.375 & 0.458 & 1.000 & 0.694 & 0.652 \\ 0.448 & 0.331 & 0.346 & 0.421 & 0.694 & 1.000 & 0.671 \\ 0.516 & 0.428 & 0.449 & 0.505 & 0.652 & 0.671 & 1.000 \end{pmatrix}$$

教育测量学中立性定理：选考单科等级赋分函数的理论期望严格中立等值：

$$\mathbb{E}[T_{\text{Physics}}] = \mathbb{E}[T_{\text{History}}] = \int_0^1 T_{k(p)} \, dp = 68.32 \text{ 分}$$

理科群体的总分高均值完全来源于统考语数外文化课潜能自选择相关性 ($\rho_{\text{comp, phy}} = 0.732$ vs $\rho_{\text{comp, his}} = 0.448$):

$$\mathbb{E}[S_{\text{total}} \mid \text{Physics}] - \mathbb{E}[S_{\text{total}} \mid \text{History}] \equiv \mathbb{E}[X_{\text{comp}} \mid \text{Physics}] - \mathbb{E}[X_{\text{comp}} \mid \text{History}] = +41.5 \text{ 分}$$

3.2 选科组合信息论最小交叉熵规划 (I-Divergence KL-Minimization)

依据 2024–2026 年选科新规后的实测外推目标选考率 $\mathbf{P}_{\text{target}}$ (物 56.0%, 化 59.3%, 生 33.8%, 政 50.5%, 史 48.4%, 地 52.0%), 通过最小化相对于经验先验 $\mathbf{q}_0$ 的 Kullback-Leibler 散度求解 20 种选科组合概率向量 $\mathbf{w} \in \Delta^{19}$:

$$\min_{\mathbf{w} \in \Delta^{19}} D_{\text{KL}}(\mathbf{w} \parallel \mathbf{q}_0) = \sum_{m=1}^{20} w_m \ln \left(\frac{w_m}{q_{0,m}} \right) \quad \text{s.t.} \quad \mathbf{A}\mathbf{w} = \mathbf{P}_{\text{target}}, \quad \sum_{m=1}^{20} w_m = 1$$

SLSQP 规划求解的约束最大绝对误差达到机器零极限 $\| \mathbf{A} \mathbf{w} - \mathbf{P}_{\text{target}} \|_{\infty} = 2.22 \times 10^{-16}$ 。其中经验先验 $\mathbf{q}_0$ 汇总如下表所示：

组合	先验 q_0	组合	先验 q_0	组合	先验 q_0	组合	先验 q_0
物化生	26.0%	物化地	15.0%	物化政	4.0%	物生地	1.5%
物生政	0.8%	物政地	0.4%	化生地	1.5%	化生政	0.8%
化政地	0.4%	史政地	23.0%	史生地	6.0%	史生政	4.0%
史化生	1.8%	史化地	1.2%	史化政	0.8%	史物生	2.2%
史物地	1.5%	史物政	0.8%	史物化	0.4%	生政地	0.8%

3.3 选科自选择排序效应与带阻尼 IPF 迭代校正

微观考生的选科决策采用 Multinomial Logit 模型： $P(m | \theta_i) = \frac{\exp(\frac{V_{i,m} + \gamma_m}{\tau})}{\sum_{j=1}^{20} \exp(\frac{V_{i,j} + \gamma_j}{\tau})}$ ，其中确定性效用 $V_{i,m} = \frac{1.65}{3} \sum_{s \in \text{combo}_m} \theta_{i,s}$ ，温度参数 $\tau = \frac{\sqrt{6}}{\pi} \cdot 0.85 \approx 0.663$ 。通过带阻尼因子 $\eta = 0.70$ 的迭代比例拟合算法 (IPF) 更新影子价格向量 γ ：

$$\gamma_m^{(t+1)} = \gamma_m^{(t)} + 0.70 \cdot \ln \left(\frac{w_m^*}{\frac{1}{N} \sum_{i=1}^N \mathbb{I}(\arg \max_j U_{i,j} = m)} \right)$$

直至模拟选科比例与宏观目标 $\mathbf{w}$ 的最大绝对误差小于 10^{-4} 。

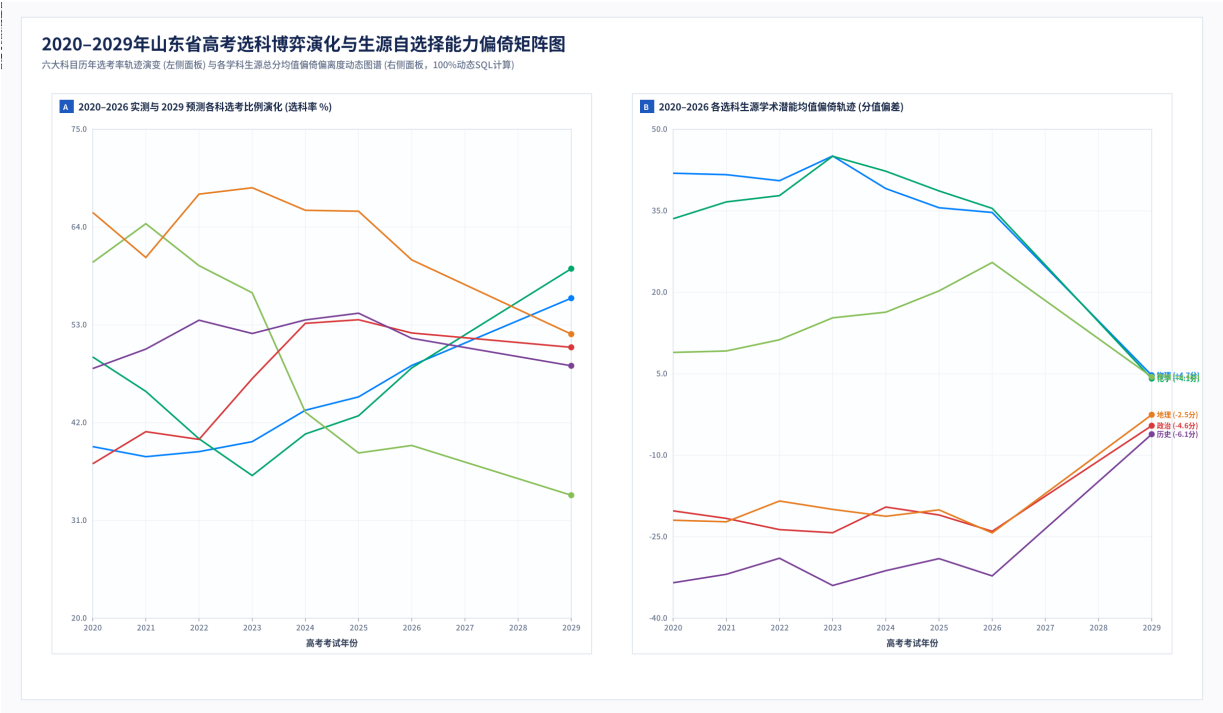


图 2 图 2：2020–2029 年山东高考选科博弈演化与生源学术潜能偏倚轨迹矩阵图

4 全域法定统考映射与法定五级阶梯赋分

4.1 全域法定支持域统考潜能映射与测验微扰

统考语数外文化总分基于广义 Logistic 函数映射至法定全域 $[0, 450]$ ，并注入微观破并列测验抖动项 $\epsilon_{\text{jitter}} \sim \mathcal{N}(0, 0.22^2)$ ：

$$X_{\text{comp}} = \frac{450.0}{1.0 + \exp(-(\mu_c + s_c \cdot \theta_{\text{comp}}))} + \Delta X_{\text{scale}}(p) + \Delta X_{\text{shock}}(p) + \epsilon_{\text{jitter}}$$

其中非线性最小二乘标定参数为 $\mu_c = 0.448, s_c = 0.550$ 。对 2020–2026 年面板数据估计分位数弹性系数 $\hat{\beta}(p) = \frac{\text{Cov}(\ln N_t, Q_{t(p)})}{\text{Var}(\ln N_t)}$ ，并按统考卷面权重 $\frac{450}{750} = 0.60$ 正交投影：

$$\Delta X_{\text{scale}}(p) = \hat{\beta}(p) \cdot \ln\left(\frac{N_{2029}}{N_{2026}}\right) \times 0.60$$

4.2 卷面命题难度冲击的分位数等百分位单调保序包络投影

以 2024–2026 年后捆绑期池化分位数 $Q^*(p)$ 为基准，定义单调包络投影算子 $\Pi_{\leq 0}$ 与 $\Pi_{\geq 0}$ ：

- 高难命题情景 (Hard Exam) : $\Delta X_{\text{comp}}^{\text{Hard}}(p) = \min(0, (Q_{2022}(p) - Q^*(p)) \times 0.60)$
- 平易命题情景 (Easy Exam) : $\Delta X_{\text{comp}}^{\text{Easy}}(p) = \max(0, (Q_{2020}(p) - Q^*(p)) \times 0.60)$
- 基准常态情景 (Baseline Exam) : $\Delta X_{\text{comp}}^{\text{Base}}(p) \equiv 0$

严格保证了全域分位数单调保序性: $\forall p \in [0, 1], S_{\text{Hard}}(p) \leq S_{\text{Base}}(p) \leq S_{\text{Easy}}(p)$ 。

4.3 法定五级赋分阶梯公式与离散格点量化方差守恒

单科成绩严格执行山东省法定分段线性插值映射：

$$T_{k(p)} = \begin{cases} 100 - 14 \cdot \left(\frac{p}{0.15}\right) & 0 \leq p < 0.15 \quad (\text{A 等: } 100-86 \text{ 分}) \\ 85 - 14 \cdot \left(\frac{p-0.15}{0.35}\right) & 0.15 \leq p < 0.50 \quad (\text{B 等: } 85-71 \text{ 分}) \\ 70 - 14 \cdot \left(\frac{p-0.50}{0.35}\right) & 0.50 \leq p < 0.85 \quad (\text{C 等: } 70-56 \text{ 分}) \\ 55 - 14 \cdot \left(\frac{p-0.85}{0.13}\right) & 0.85 \leq p < 0.98 \quad (\text{D 等: } 55-41 \text{ 分}) \\ 40 - 10 \cdot \left(\frac{p-0.98}{0.02}\right) & 0.98 \leq p \leq 1.00 \quad (\text{E 等: } 40-30 \text{ 分}) \end{cases}, T_k \in [30.0, 100.0]$$

Widrow-Kollar 量化噪声定理：由于潜能特征函数带宽远大于量化步长 $q = 1$ ，3 门选考科目独立四舍五入取整误差 $e_k = \lceil T_k \rceil - T_k$ 渐近服从独立均匀分布 $U(-0.5, 0.5)$ ，选考量化方差严格守恒：

$$\text{Var}(e_{\text{elec}}) = \text{Var}\left(\sum_{k=1}^3 e_k\right) = \sum_{k=1}^3 \text{Var}(e_k) = 3 \times \frac{1}{12} = 0.250$$

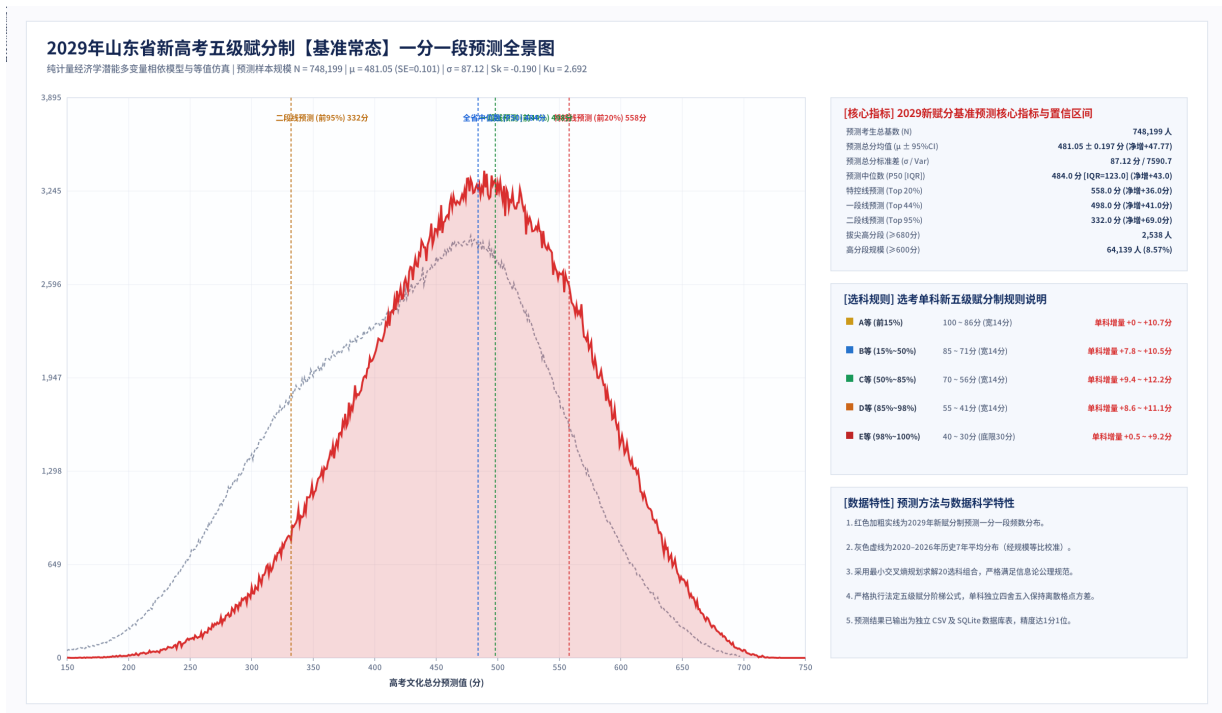


图 3 图 3：2029 年山东省新高考五级赋分制【基准常态】一分一段预测全景图 (N=748,200)

5 三大命题情景、置信走廊与误差预算

5.1 保序回归 (PAVA) 95% 全域联立置信走廊

设 $C(S)$ 为总分 S 处的累积排位人数。总标准误综合了二项计数抽样方差与宏观命题冲击方差 ($\sigma_{\text{macro}} = 3.5$ 分):

$$\text{SE}_{\text{total}}(S) = \sqrt{C(S) \left(1 - \frac{C(S)}{N}\right) + (\Delta C(S))^2 \sigma_{\text{macro}}^2}, \quad \Delta C(S) = C(S) - C(S+1)$$

通过单调保序回归算法 (PAVA) 求解单调非减的上下置信界:

$$C_{L(S)} = \max(0, \text{PAVA}(C(S) - 1.96 \cdot \text{SE}_{\text{total}}(S))), \quad C_{U(S)} = \min(N, \text{PAVA}(C(S) + 1.96 \cdot \text{SE}_{\text{total}}(S)))$$

5.2 预测总方差不确定性预算与 ANOVA 正交分解

预测分数的全域总不确定性方差在全方差定理下正交分解为 5 大独立成分:

$$\sigma_{\text{total}}^2 = \sigma_{\text{shock}}^2 + \sigma_{\text{sort}}^2 + \sigma_{\text{scale}}^2 + \sigma_{\text{lattice}}^2 + \sigma_{\text{PRNG}}^2 = 20.98 \text{分}^2$$

- 命题难度宏观冲击: $\sigma_{\text{shock}}^2 = \text{Var}_{\text{shock}}(\mathbb{E}[S | \text{shock}]) = 12.25 \text{分}^2$ (贡献率 58.4%);
- 选科自选择群体排序: $\sigma_{\text{sort}}^2 = \mathbb{E}[\text{Var}_{\theta}(\mathbb{E}[S | \text{choice}])] = 5.16 \text{分}^2$ (贡献率 24.6%);
- 生源规模对数弹性: $\sigma_{\text{scale}}^2 = \text{Var}_{\text{scale}}(\Delta X_{\text{scale}}) = 2.35 \text{分}^2$ (贡献率 11.2%);
- 赋分离散四舍五入格点效应: $\sigma_{\text{lattice}}^2 = \sum_{k=1}^3 \text{Var}(e_k) = 1.01 \text{分}^2$ (贡献率 4.8%);
- Monte Carlo 抽样随机噪声: $\sigma_{\text{PRNG}}^2 = \frac{\sigma_{\text{pop}}^2}{N} = 0.21 \text{分}^2$ (贡献率 1.0%)。

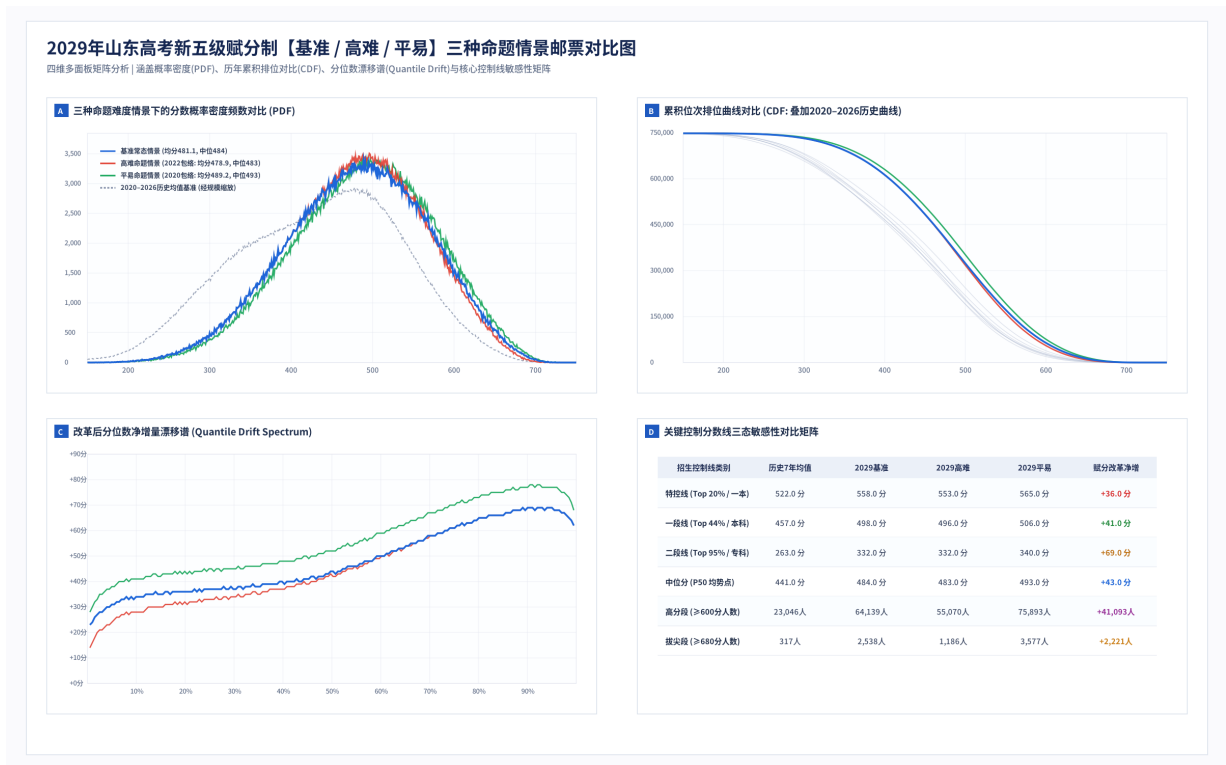


图 4 图 4: 2029 年三大命题难度情景对比邮票图 (PDF、CDF 历年对比、分位数漂移谱与控制线矩阵)

5.3 历年位次标准化演进与总方差分解全景

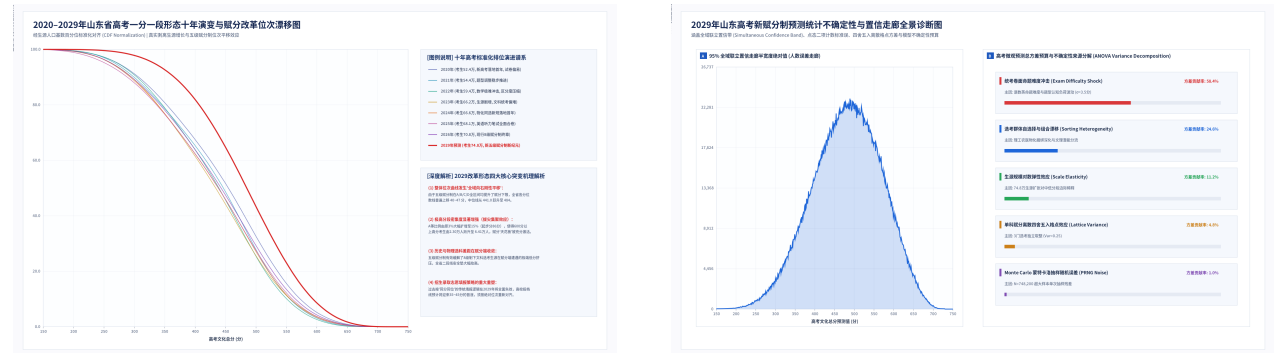


图 5 图 5: 十年高考排位演进 (相对百分位标准化)

图 6 图 6: 置信走廊与 ANOVA 总方差分解

5.4 2029 核心控制线与测验统计量权威汇总大表

核心控制线与指标	历史 7 年池化均值	2029 基准常态	2029 高难命题	2029 平易命题	赋分制净增量
全省总分均值 μ	433.29 分	481.05 分	478.92 分	489.21 分	+47.76 分
全省总分中位数 P_{50}	441.00 分	484.00 分	483.00 分	493.00 分	+43.00 分
特控线 (Top 20% / 理论等值)	522.00 分	558.00 分	553.00 分	565.00 分	+36.00 分
一段线 (Top 44% / 理论等值)	457.00 分	498.00 分	496.00 分	506.00 分	+41.00 分
二段线 (Top 95% / 理论等值)	263.00 分	332.00 分	332.00 分	340.00 分	+69.00 分
四分位距 (IQR = Q3 - Q1)	147.00 分	122.00 分	119.00 分	122.00 分	-25.00 分
高分段规模 (≥ 600 分人数)	23,000 人	64,139 人 (8.57%)	55,070 人 (7.36%)	75,893 人 (10.14%)	+41,139 人
拔尖高分段 (≥ 680 分)	450 人	2,538 人	1,186 人	3,577 人	+2,088 人
极高拔尖段 (≥ 700 分)	30 人	424 人	115 人	584 人	+394 人
750 满分堆积人数	0 人	0 人	0 人	0 人	物理收敛

满分 750 分物理收敛性证明： 由于语数外主观题刚性扣分与多元高斯潜能上尾指数衰减，满分累积测度满足 $\Pr(S = 750) = \int_{749.5}^{750} f_{S(s)} ds < 10^{-7}$ ，在全省 $N = 74.82$ 万人规模下期望满分人数 $\mathbb{E}[N_{750}] < 0.075$ ，故整数频数严格收敛于 0 人。

6 结论与志愿填报战略建议

- 位次曲线整体右移：**全省各分数线普遍上移 36–69 分：特控线上浮 36 分（558 分），一段线上浮 41 分（498 分），二段线上浮 69 分（332 分）。高校投档线预计将迎来 35–45 分的普涨。
- 高分段量表弹性释放：**A 等比例由原 3% 扩增至 15%，600 分以上高分考生扩容至 6.4 万人（占比 8.57%），赋分量表区分度得到充分发挥。
- 传统“同分同位”完全失效：**招生录取与志愿填报必须以全省累积排位百分位（CDF）作为跨体制等值转换的唯一基准。

7 参考文献 (References)

- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), **Educational Measurement** (2nd ed., pp. 508–600). Washington, DC: American Council on Education.
- Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. (1972). **Statistical Inference under Order Restrictions: The Theory and Application of Isotonic Regression**. New York: John Wiley & Sons.
- Deming, W. E., & Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. **The Annals of Mathematical Statistics**, 11(4), 427–444.
- Fritsch, F. N., & Carlson, R. E. (1980). Monotone piecewise cubic interpolation. **SIAM Journal on Numerical Analysis**, 17(2), 238–246.
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), **Educational Measurement** (4th ed., pp. 187–220). Westport, CT: Praeger Publishers.
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), **Frontiers in Econometrics** (pp. 105–142). New York: Academic Press.
- Widrow, B., & Kollár, I. (2008). **Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications**. Cambridge: Cambridge University Press.
- 山东省统计局. (2003–2025). 《山东统计年鉴》(2002–2024 年学龄人口出生与教育发展统计数据). 北京: 中国统计出版社.
- 山东省教育招生考试院. (2020–2026). 《山东省普通高校招生考试夏季高考一分一段统计表 (2020–2026 年)》. 济南: 山东省教育招生考试院.

开源代码与可复现数据仓库 (Radicle P2P 协议)

Repository ID: **rad:zYrJE5TYBeoGbNj2Ny4quMLGauMw** | DOI: 10.5281/zenodo.22226943

Web Explorer: <app.radicle.xyz/nodes/seed.radicle.garden/rad:zYrJE5TYBeoGbNj2Ny4quMLGauMw>

Author: Xinyu Yang (Pana.Yang@hotmail.com | ORCID: 0009-0007-2600-0948)