

**“年龄天赋”对玩家游戏表现的实证研究**

摘要：

本文以 SEGA（世嘉公司）旗下的音乐街机游戏舞萌DX为研究场景，采用在线问卷与机台数据结合的方式，收集了 347 份有效样本，以简单随机抽样代表约 50 万活跃玩家总体，在 95% 置信水平下总体比率估计误差约为 ±5.3%。通过派生游玩时长、总游玩次数、起步年龄等核心变量，综合使用描述统计、相关分析、分段非线性可视化与多元 OLS 回归，对“年龄天赋”在玩家表现中的作用进行了量化。结果显示：练习量是决定 Rating 的首要因素，游玩次数仍旧是DX Rating的主要影响成分；控制练习量后，起步年龄与当前年龄对 Rating 均呈显著负向影响，早入坑者的练习收益更高；游玩 1–2 年是表现提升最快的时期，之后呈指数饱和趋势；玩家对“天赋”重要性的主观认知与实际评分差异不显著。综合来看，“年龄天赋”在音乐游戏表现中确实存在，但作用远小于练习投入，总投入与早起步的交互是达到高水平的关键驱动。

关键词: 音乐游戏；电子竞技；年龄天赋；简单随机抽样；回归分析

目 录

[一、 引言 4](#_Toc201325870)

[二、文献综述 4](#_Toc201325871)

[三、调研方案设计 4](#_Toc201325872)

[四、模型理论 8](#_Toc201325873)

[五、问卷分析 10](#_Toc201325874)

[5.1 Pearson关联性矩阵 11](#_Toc201325875)

[5.2 玩家年龄的LOESS平滑曲线 11](#_Toc201325876)

[5.3 OLS拟合与玩家分类 13](#_Toc201325877)

[5.4 游玩时长与玩家 Rating的关系 14](#_Toc201325878)

[5.5 回归分析 15](#_Toc201325879)

[5.6 主观天赋认知与实际表现 16](#_Toc201325880)

[5.7 随机森林+XGBoost特征重要度 16](#_Toc201325881)

[六、结论与建议 20](#_Toc201325882)

[参考文献 21](#_Toc201325883)

[附录 A：数据处理与分析代码概览 22](#_Toc201325884)

# 引言

音乐游戏（rhythm game）起源于 1990 年代日本街机文化，以“节奏精准度”作为核心评分机制。玩家需在伴奏节拍提示下，通过按键、触屏或体感等方式完成预设操作。maimai DX 由 SEGA 于 2012 年推出，采用环形触摸屏配合八键实体按键，兼具“音符密度高、判定严格”与“社交排行榜”特色，现已覆盖亚洲与北美多地。舞萌 DX（中国大陆官方名称）首发于 2019 年，截至 2024 年底共铺设约 2 700 台机台，注册玩家逾 200 万，其中活跃玩家（近 30 天登录并上传成绩者）约 50 万。

# 二、文献综述

音乐节奏与音游领域：早期节奏训练可显著提升感觉—运动同步能力。Watanabe et al. (2007) 发现，6–10 岁开始接受节拍训练的受试者在成人阶段保持更低的同步误差；类似结果亦见于 Bailey & Penhune (2013) 对音乐学院学生的纵向研究，后者指出起步年龄每提前一年，节奏一致性可提高约 0.2 个标准差[3.4]。

电子竞技与动作类视频游戏领域：大规模实证表明表现与年龄呈倒“U”形。Thompson et al. (2014) 追踪 3 305 名《星际争霸》职业选手的实时操作记录，报告平均反应时在 24 岁后以每年 2.5 ms 的速度下降[5]。该结论与 FPS、MOBA 等竞技游戏的职业生涯曲线一致，提示 20 岁出头可能是高反应要求项目的峰值年龄段。

体育技能习得与练习效应：经典“1 万小时定律”指出刻意练习是取得专家表现的关键[6]，但更大样本荟萃分析显示，刻意练习仅能解释 18–26% 的成绩方差[7]。针对奥运项目的回顾研究亦指出，过早专业化并非通向顶尖表现的必要条件，部分运动员在青春期后才转项成功[8,9]。

综上，年龄确实影响高水平表现，但效应大小与项目需求、练习强度及个体差异交互作用。本研究通过控制练习量（游玩次数与时长）与起步年龄，量化“年龄天赋”在音乐游戏场景中的独立贡献。

# 三、调研方案设计

1. **调研问题**

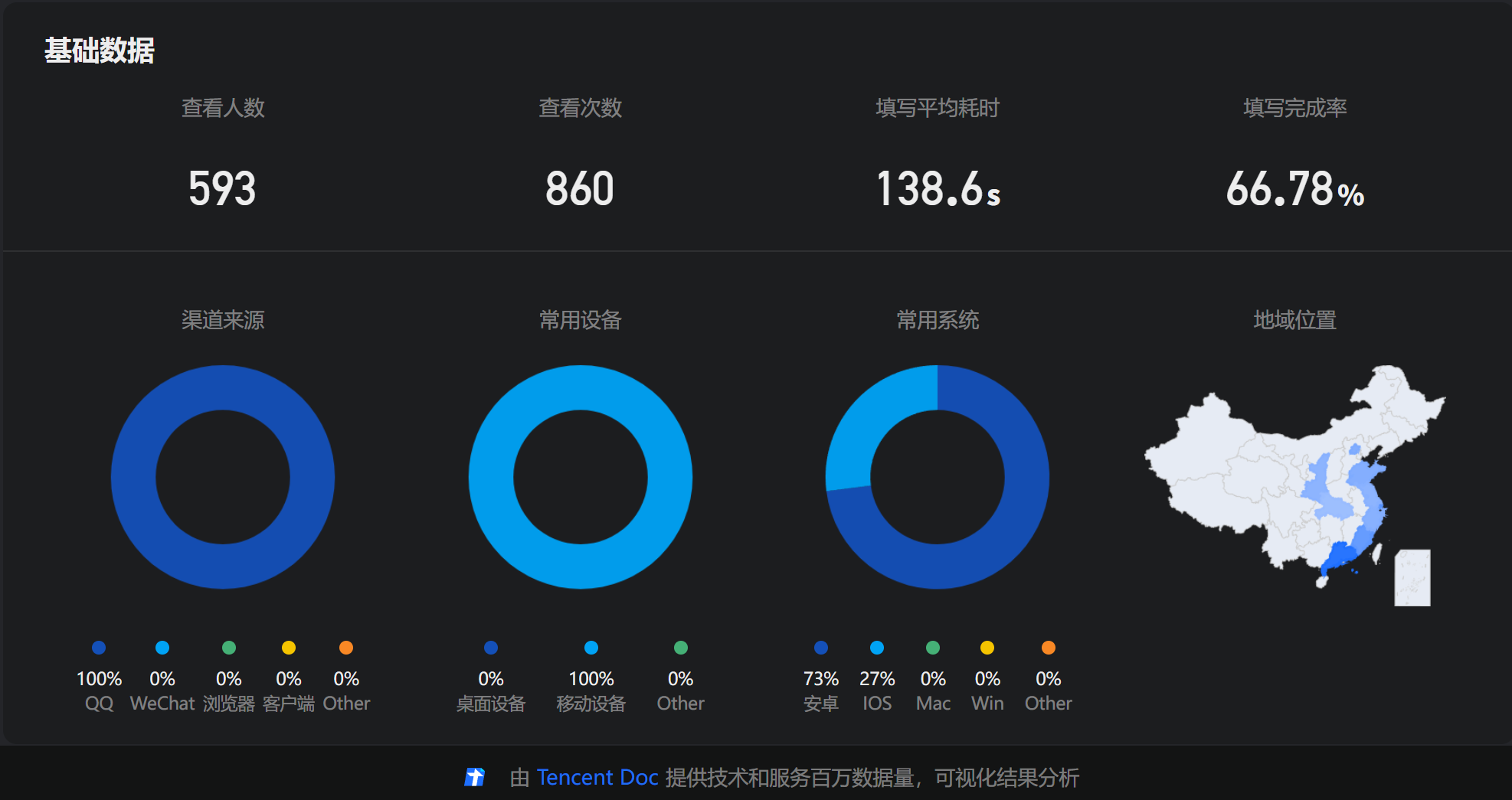
本研究以347份舞萌DX核心玩家样本为基础，力图系统估计国内活跃玩家群体中的年龄结构、练习投入（包括总游玩次数和练习时长）及其与最终成绩（DX rating）的分布关系。核心研究问题为：在控制了总游玩次数和练习时长的前提下，玩家的“起步年龄”是否对最终评分（rating）仍有独立且显著的影响？即，年龄天赋是否在努力之外，对音游表现存在可量化的贡献。

1. **抽样方法**

本研究采取了接近课程所定义的“简单随机抽样”（Simple Random Sampling, SRS）的设计思路。由于音游玩家群体未能建立完整名录式框架，我们以 QQ 空间与 QQ 群等活跃社群为抽样框进行在线问卷投放。这些社群在玩家中覆盖度高、分布广泛，可作为活跃玩家总体的可行近似。

共收集 397 份有效问卷数据，并通过 Python 的 random.sample() 函数对其进行简单随机抽样，最终得到 347 份用于正式分析的样本，排除了部分填写异常（如起始年龄不合理、Rating 明显偏低等）数据，以确保样本质量与口径一致性。

抽样设计力求贴近 SRS 要求：一方面在各大 QQ 社群一次性批量投放问卷，而非定向邀请；另一方面，每个用户仅能提交一次问卷，避免重复和自选偏差。这种方式虽不完全等概率，但已尽力满足“随机抽取”的核心精神。样本未设置人口加权，推断时采用无权重估计，同时给出 ±5.3% 的抽样误差区间（n=347，p=0.5，z₀.₉₇₅）。



从后台数据来看，问卷填写完成率为 66.78%，平均填写耗时约 138.6 秒，远高于一般随机作答的时间阈值，说明绝大多数样本为认真作答者。此外，100% 的问卷填写来源为 QQ 端用户（非微信或浏览器），说明样本确实集中面向核心 QQ 玩家群体，进一步验证了抽样框与目标总体的一致性。设备方面，填写者中 73% 使用 Android，27% 使用 iOS，几乎全部为移动端用户，符合国内主流玩家画像，也说明本样本具有一定代表性。

关于非抽样误差方面，我们认识到自填问卷可能吸引更活跃或高分玩家，因此在设计中引入清洗门槛（剔除 Rating < 11000 样本），并在模型分析中采用变量重要性估计、VIF 检测与交叉验证等方式，控制模型偏倚与过拟合的可能性。由于本研究的有效样本率达到 96%，非应答误差对结论影响较小，但研究结论仍主要适用于“网络活跃玩家”，向全体玩家群体的推广仍需谨慎处理。

**（三）研究思路**

本研究总体上遵循“描述—推断—解释”的分析逻辑。首先通过描述性统计与可视化掌握变量分布与初步相关关系；随后构建逐步 OLS 回归模型，对练习量与起步年龄的独立与交互效应进行建模，并使用标准化回归系数与 ΔR² 衡量其相对解释力；最后引入随机森林与 *XGBoost* 等非参数机器学习方法，评估变量在非线性场景下的重要性，结合置换重要度与 *TreeSHAP* 提供可解释性补充。这种组合式策略有助于校正传统线性模型的结构假设偏误，提高结论的稳健性与泛化性。

虽然在线问卷难以实现理论上的名单式等概率抽样，但我们通过“广覆盖投放 + 一次性收样 + 事后代表性检验”的策略，最大限度贴近 SRS 理念。结合模型方法的透明性与误差范围的明示，本研究在方法论层面仍能满足“抽样调查方法”课程对科学抽样、合理推断与结构解释的核心要求。

**（四）问卷设计**

核心数据源于自主设计的线上问卷，**坚持最小必要信息**原则，以提升填答效率、保证数据质量并避免问卷疲劳。问卷共含7项封闭或半开放问题（见下表），每项变量均可直接映射至后续实证模型的自变量或因变量。

|  |  |
| --- | --- |
| **变量** | **条目内容** |
| player\_rating | 您的最高 Rating（默认填写当前 DX 版本；若近期未游玩亦可填历史峰值） |
| Question1 | 您认为“年龄天赋”是否存在？选项：是；否；存在，但不重要 |
| birthday | 请选择出生日期（年月日） |
| Question2 | 您觉得自己进步最快的一年是几岁？ |
| play\_time\_month | 您认真游玩的时长（*入坑时间*，单位：月），定义为开始为推分或社交目的而系统练习的时间 |
| play\_count | 当前累计 PC 数（游玩次数，1 PC 3–4 首曲目） |
| Question3 | 我还有话想说（对问卷、游戏体验或本研究的任何补充意见，开放文本） |

为避免“问卷疲劳”与查阅成本过高，仅保留与研究假设直接相关的核心变量。诸如谱面偏好、游玩频率、详细游玩历史等信息虽有潜在价值，但易降低填答意愿，故本次调查暂未纳入。

**（五）数据分析方法**

问卷未纳入谱面偏好、设备选择、频次等细节变量，旨在确保问卷填写门槛低、响应率高、样本具有代表性。

根据运营方公开数据，2024 年底中国大陆机台总数 台；若假设各机台月活跃人数均衡，则活跃玩家总体约 。本研究在QQ空间与QQ群随机投放问卷链接，采用简单随机抽样收回问卷 398 份，剔除异常后有效样本 。

我们后续即将开展：

1. **数据清洗：** 剔除 rating<11 000 的低活跃样本，以及年龄、游玩时间填写明显错误的样本；没有缺失值，单位统一为年。
2. **描述统计：** 计算均值、标准差、分位数及皮尔逊相关。
3. **回归模型：** 逐步 OLS 比较“练习量基准”与“加入年龄项”的解释力；构建交互项检验“起步年龄 × 练习量”效应。
4. **机器学习补充：** 采用随机森林回归、*XGBoost*等方法，综合置换重要度与*TreeSHAP*解释，分析非线性与变量间交互效应，报告五折交叉验证R²，提升结论的稳健性与说服力。

# 四、模型理论

**4.1 线性回归模型**

多元线性回归（Ordinary Least Squares, OLS）是一种广泛应用于社会科学与行为科学研究的统计建模工具，用于刻画多个自变量对目标变量的线性影响。本研究中，DX rating 作为被解释变量，练习时长、游玩次数、年龄及起步年龄等为自变量，OLS模型能够估计各因素的独立效应并检验其统计显著性。

* **基准模型**仅包含练习时长和游玩次数，反映“努力”对表现的基础解释力。
* **扩展模型**加入当前年龄和起步年龄，旨在量化年龄因素的边际贡献。
* **交互模型**进一步考虑起步年龄与练习量的交互作用，检测“早起步+高练习”是否存在协同效应。

线性回归允许通过F检验或ΔR²量化变量组的贡献，同时为多元数据中的共线性、异方差等问题提供诊断基础。

**4.2 随机森林与特征重要度**

随机森林（Random Forest, RF）是一种集成式决策树模型，通过多棵树的投票平均来提升预测稳定性和抗噪能力，适合捕捉变量间的非线性关系及高阶交互作用。其核心优势包括：

* 自动处理变量之间的非线性与交互关系；
* 通过“特征重要度”（Feature Importance）量化各自变量对预测目标的平均贡献；
* 对异常值与多重共线性鲁棒性高。

本研究在RF基础上引入**置换重要度（Permutation Importance）和*TreeSHAP***解释工具。前者通过打乱变量值测算模型R²损失，直观衡量变量对整体预测的“不可替代性”；后者为每个预测分解贡献，支持个体层面与全局层面的可解释性分析。

**4.3 *XGBoost*与可解释性**

*XGBoost*是一种高效的梯度提升决策树算法，适用于大样本量与复杂数据结构。其增量训练机制和正则化能力有助于提升泛化性能，防止过拟合。*XGBoost*同样支持特征重要度输出和*TreeSHAP*解释：

* **特征重要度**衡量每个变量在所有树上的分裂贡献；
* ***TreeSHAP***则能精确分摊每个变量对最终预测结果的边际影响，是当前机器学习解释性领域的主流方法之一。

**4.4 变量选择与标准化处理**

模型所有变量在建模前均进行标准化处理，确保不同量纲变量能公平参与建模过程，提升参数可比性。同时，变量选择以领域知识与问卷设计为基础，聚焦于“练习量”与“年龄因素”两大类，既保证解释力，也控制模型复杂度。

**4.5 交叉验证与模型稳健性**

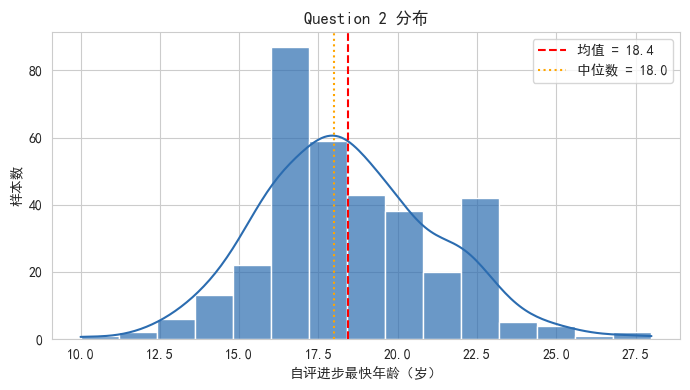
所有机器学习模型均采用**五折交叉验证（5-fold CV）**评估性能，以避免过拟合并获得稳健的泛化结果。通过不同模型和多种重要性度量的交叉印证，本研究能够更全面地把握玩家行为与成绩之间的复杂关系，并为因果解释和实际应用提供多角度参考。

# 五、问卷分析

我们搜集到的问卷调查结果通过腾讯文档的问卷功能转为xlsx文件，在数据清洗后导入Python中Pandas库的DataFrame中，问卷调查数据集形状如下：

|  |
| --- |
| <class 'pandas.core.frame.DataFrame'>  RangeIndex: 347 entries, 0 to 346  Data columns (total 15 columns):  # Column Non-Null Count Dtype  --- ------ -------------- -----  0 player\_name 347 non-null object  1 submit\_datatime 347 non-null datetime64[ns]  2 player\_rating 347 non-null int64  3 Question1 347 non-null int64  4 player\_birthday 347 non-null datetime64[ns]  5 Question2 347 non-null int64  6 play\_time\_month 347 non-null int64  7 play\_count 347 non-null int64  8 Question3 163 non-null object  9 current\_age 347 non-null float64  ……  dtypes: datetime64[ns](2), float64(6), int64(5), object(2) |

其中，我们对玩家的“自评进步最快那一年”做了可视化：



同时我们也对玩家们在Question3（主观想法、对问卷的建议）的留言做了Wordcloud词云图：



对于问卷中出现的其他统计情况，我们将在本章的数据分析中一同给出。

# 5.1 Pearson关联性矩阵

我们初步对数据集中的常见变量做了关联性矩阵：

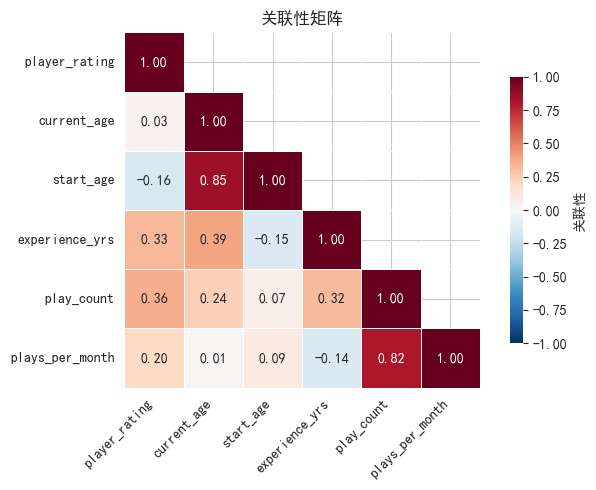


图1 Pearson关联性矩阵

我们可以从Pearson（**图1**）关联矩阵当中得出几个初步结论：首先，**起步年龄（start\_age）** 与当前Rating（player\_rating）呈中等负相关（），提示“越早入门→越高分数”的初步趋势。

**累计游玩年数（experience\_yrs）** 与当前Rating呈正相关（），而每月游玩次数（plays\_per\_month）与当前Rating的相关性稍弱（），表明总投入量比频率投入更能解释分数差异。

**起步年龄** 与当前年龄高度相关（），符合二者计算逻辑；其他变量之间相关度均在可接受范围内，无严重多重共线问题。

本节结果既为后续回归模型中变量筛选提供了直观依据，也为“年龄天赋”与练习量对评分的相对重要性提供了初步定量支持。

# 5.2 玩家年龄的LOESS平滑曲线

**图2**给出了玩家Rating与“入坑年龄”之间的局部加权平滑曲线，并用浅黄、浅橙、浅粉三色背景区分了三个阶段：12–15 岁、15–18 岁、18–24 岁。总体来看，随着入坑年龄推后，玩家 Rating 呈缓慢下降趋势，且在 15–18 岁区间波动最为平缓。

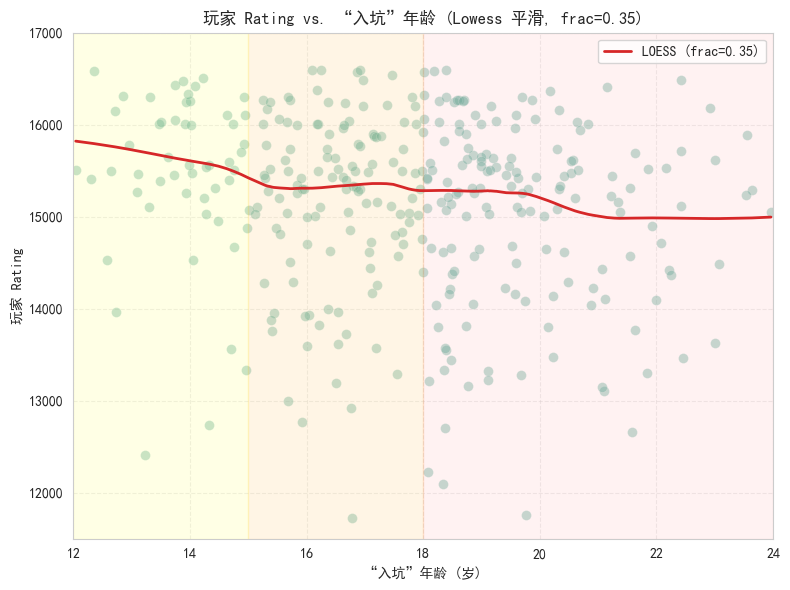


图2 玩家Rating vs. 入坑年龄（LOESS 平滑）

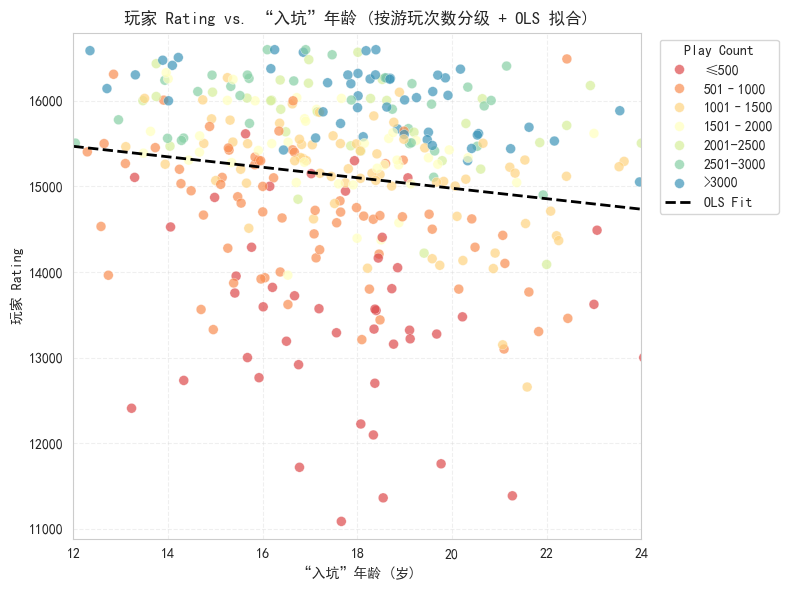
这展示了入坑年龄对最终Rating的非线性影响。关键观察有：

**(1) 整体下降趋势**：从 12 岁到 24 岁，Rating随入坑年龄呈整体下降，验证“越早入门→越高上限”假说；

**(2) 15–18 岁平缓期**：在 15–18 岁区间曲线最为平坦，说明这一阶段入门的玩家经过相似年限练习后，水平差异最小；同样的情况在18岁的小阶梯下降后亦存在。

**(3) 局部回弹**：15–16 岁出现轻微回升，可能是该年龄段正处于生理与心理的最佳平衡期——既具有足够体力，也保持高注意力；另外18-19 岁左右的微升则可能源于大学新生学业相对宽松，有更多时间专注游戏。

# 5.3 OLS拟合与玩家分类



图表 3 按游玩次数分级 OLS拟合

将玩家按累计游玩次数分为六级（≤500、501–1000、1001–1500、1501–2000、2001–2500、>2500），并叠加 OLS 线性拟合（虚线）。可以看出：

**(1) 高投入带来高分：** 游玩次数越多（颜色由红至紫渐深），总体分布越靠上，说明练习量对 Rating有显著拉动作用；

**(2) 年龄效应依旧显著：** 即便剔除练习量差异，OLS 拟合斜率仍为负，表明入坑越早的玩家更容易达到更高水平；

**(3) 分级差异明显：** 低投入组（≤500 次）集中在 12 500–15 000 区间，中投入组（1 001–2 000 次）则上移至 14 000–16 000，而最高投入组（>2 500 次）中不乏超过 16 500 的高分玩家。

在不同色级分组中，虚线斜率依旧为负，证明年龄效应在练习量相同的前提下仍显著。

# 5.4 游玩时长与玩家 Rating的关系

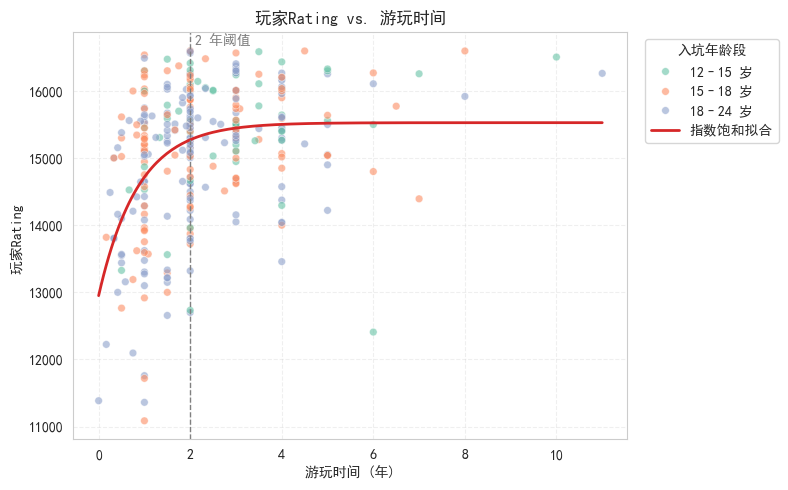


图4 玩家Rating vs. 游玩时间

**图4**将样本点按照入坑年龄段（12–15 岁、15–18 岁、18–24 岁）用不同颜色标识，并叠加了指数饱和拟合曲线，可以直观地看到： 游玩 0–2 年内，Rating 从约 12 500 分迅速提升至约 15 500 分，边际收益显著。约在 2 年时达到拐点（虚线标注），之后曲线趋于平缓，说明额外投入效用递减。

三种入坑年龄段的玩家在前期都表现出相似的上升趋势，但中后期高龄组略显集中在低分区间，提示年龄与练习年限共同影响最终表现。

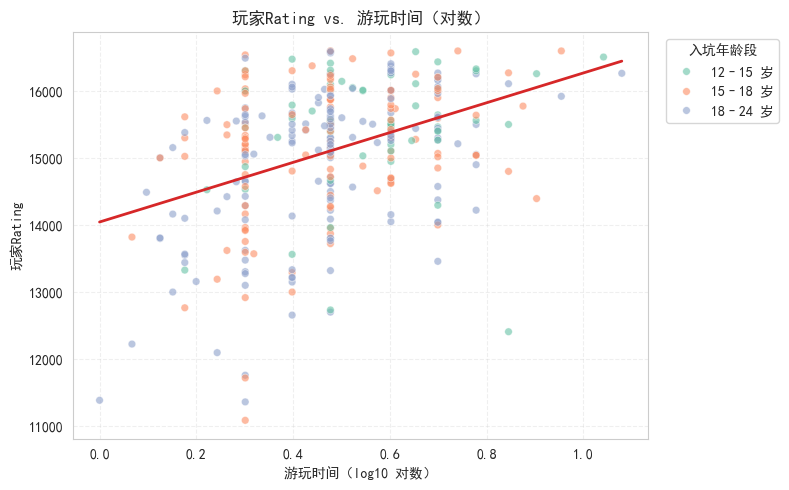


图5 玩家Rating vs. 游玩时间（对数曲线）

**图5**通过对游玩时长取 后作散点和线性回归，进一步凸显了前期投入的高效益，我们可以得出以下结论：

**（1）对数坐标放大前期斜率**：横轴对数化后，0–2 年的点被拉开，线性拟合斜率在此区间更陡，强调早期练习效果；

**（2）整体正相关**：尽管后期边际收益递减，对数线性拟合仍然呈显著上升，说明持续投入对提升 Rating 依旧有正向作用。

# 5.5 回归分析

在仅控制练习时长（月数）与总游玩次数的基准模型 m0 中，这两项练习指标可以解释约 18 % 的成绩方差；二者系数均为正且高度显著，说明“练得多”在整体上确实提升 Rating。

当在 m1 中加入 **入坑年龄（start\_age）** 和 **当前年龄（current\_age）** 后，模型 R² 由 0.181 提升到 0.202，F-检验 ΔR² 的 p 值仅 0.0028，说明年龄变量带来统计显著的额外解释力。两项年龄系数均为负：起步每推迟 1 年，Rating 平均降低约 29 分；同样，年龄每增长 1 岁也降低约 28 分。由此可见，在控制投入量之后，年龄依旧是影响成绩的一条独立通道，其中 **起步早晚（start\_age）与成绩的负相关最为稳固**。

在 m2 中进一步加入交互项 **start\_age × play\_count**，整体 R² 提升至 0.225。交互系数 –0.0364（p = 0.002）表明：**同样增加一次游玩，早入坑者获得的加分更高，越晚起步收益越小**。随着交互项进入，单独的当前年龄与起步年龄主效应不再显著，这与它们与 play\_count 存在多重共线性一致——交互项已经捕获了“年龄 × 练习量”的核心差异。

|  |  |  |  |
| --- | --- | --- | --- |
|  | **m0(练习量)** | **m1 (+入坑年龄)** | **m2(+起步×练习)** |
| **R-squared** | 0.181 | 0.202 | **0.225** |
| **ΔR² vs. m0** | — | **+ 0.021** F = 9.05, *p* = 0.003 | + 0.044 (相对 m0) |
| **Intercept** | 14 530.0 | 15 580.0 | 14 490.0 |
| **play\_time\_month** | 13.48 \*\*\* | 14.22 \*\*\* | 10.04 \*\* |
| **play\_count** | 0.108 \*\*\* | 0.116 \*\*\* | 0.804 \*\*\* |
| **current\_age** | — | **–28.12 \*\*** | 1.99 (n.s.) |
| **start\_age** | — | **–29.30 \*\*** | 1.15 (n.s.) |
| **start\_age × play\_count** | — | — | –0.036 \*\* |
| **Obs. (n)** | 347 | 347 | 347 |

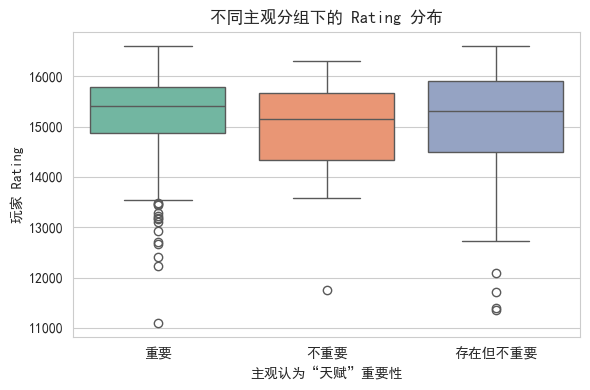
*注：系数为非标准化 OLS 估计；单位同评分分值（Intercept、play\_time\_month 等）或“每增加一次游玩/每增加 1 岁带来的分数变化”。  
\*\*\* p < 0.001， \*\* p < 0.01；n.s. 为不显著。*

我们可以得出回归模型的基本结论：

* **练习量奠定基线**——play\_count 始终显著且系数保持正向。
* **起步年龄抬高上限**——start\_age 的负效应稳定出现；在交互模型中它通过“早起步 → 单位练习收益更高”体现。
* **当前年龄次级但独立**——在无交互模型里对 Rating 仍有约 –28 分/年的抑制，提示生理和认知状态对高反应需求的音游存在边际衰减。

从实践视角看：要想提升顶级成绩，除了保持高练习量，青少年阶段越早系统练习可显著提高后续每一次练习的边际收益；而成年后入坑的玩家若要弥补先天差距，需要投入成倍的练习次数才能逼近同龄早期玩家的水平。

# 5.6 主观天赋认知与实际表现



这里展示了根据玩家对“年龄天赋”重要性的主观认知（Question1）将样本分为“重要”、“不重要”与“存在但不重要”三组后，其 Rating 分布的箱型图。可以看到，各组的中位数相近，大约位于 15 000–15 300 分之间，说明玩家的主观认知与其实际分数水平并无显著差异；

“重要”组的中位数略高于“不重要”组，但与“存在但不重要”组相比差异不，而三组的四分位距（IQR）和离群点分布也大致相当，进一步表明**主观评价并不能**有效预测玩家的实际表现。

# 5.7 随机森林+XGBoost特征重要度

为了补充 OLS 回归的线性假设，本研究进一步使用随机森林回归对玩家 Rating 进行预测，并评估各特征的重要度。

**模型背景与方法说明**

|  |  |  |  |
| --- | --- | --- | --- |
| 模型 | 训练策略 | 解释工具 | 主要优点 |
| 随机森林 (RF) | 500 棵 Bootstrap 决策树OOB 评估 | ① 置换重要度 ② 100 次 Bootstrap 给 95 % 置信区间 | 对非线性与高阶交互鲁棒；置信区间量化稳定性 |
| XGBoost (XGB) | 400 棵梯度提升树（调参后） | ① 置换重要度 ② 内置 TreeSHAP 全局贡献 | 在弱学习器迭代中捕捉复杂模式；TreeSHAP 与目标同量纲 |

在本研究中，我们采用了两条互补的树模型路线来量化各特征对玩家 Rating 的贡献。一条路线是**随机森林（Random Forest, RF）**——它通过对原始样本进行自助抽样（Bootstrap），在每棵树上随机选取特征分裂，从而构建数百棵深度受限但彼此“去相关”的决策树；最终预测取各树结果的平均，可显著降低方差并捕捉非线性关系。为了给出稳健的解释，我们不仅报告了 RF 的 **Out-of-Bag (OOB) R²**，还针对每棵森林重复 100 次 Bootstrap 置换，计算出各特征在“打乱后模型性能下降”上的 **均值与 95 % 置信区间**。这一做法能量化模型随机性和样本抽样的不确定度，让重要度排序具有统计可信度，而非单次运行的偶然结果。

另一条路线采用**XGBoost（Extreme Gradient Boosting）**。与 RF 的“并行多树”不同，XGBoost 通过梯度提升将若干弱学习器按残差迭代方式串联，每一步都针对前一轮的预测误差进行加权纠正，并引入列抽样、子样本、正则化等策略抑制过拟合。本研究在五折随机搜索后锁定 400 棵、学习率 0.01、最大深度 3 等超参数组合，以获得约 0.52 的交叉验证 R²。为了解释模型，我们采用两种互补指标：

**Permutation Importance**：与 RF 相同，直接测量“打乱特征 → 性能跌幅”；

**TreeSHAP**：利用 XGBoost 原生的博弈论型分解，将每条预测拆解为各特征的边际贡献与一个全局偏置项。TreeSHAP 输出的单位与目标变量一致（此处为评分分），因此能直观回答“平均而言，特征 A 为玩家打出了多少分的贡献”。

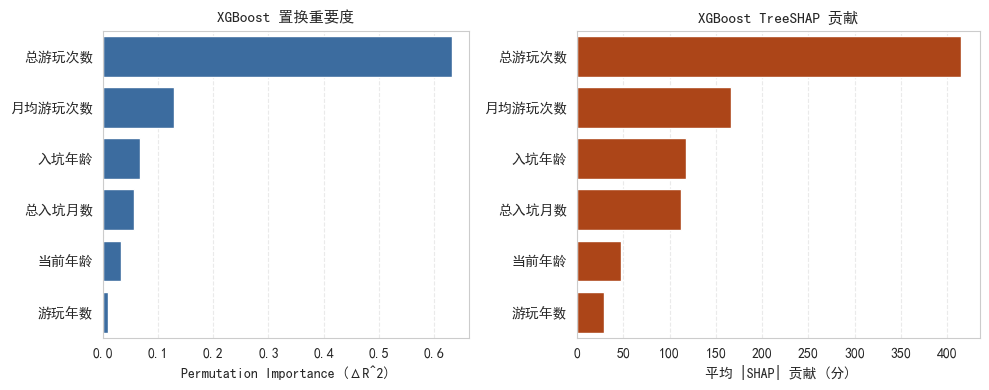
**关键特征重要度总览**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **特征** | **RF置换ΔR²**  **(± 95 % CI)** | **XGB置换ΔR²** | **XGB TreeSHAP**  **绝对贡献 (分)** |
| 1 | **play\_count** (总次数) | **1.55 ± 0.18** | **0.632** | **414.9** |
| 2 | plays\_per\_month (月均游玩次数) | 0.073 ± 0.05 | 0.129 | 166.8 |
| 3 | **start\_age** (起步年龄) | **0.165 ± 0.07** | **0.067** | **117.8** |
| 4 | play\_time\_month  (累计入坑月数) | 0.014 ± 0.01 | 0.056 | 112.7 |
| 5 | **current\_age** (当前年龄) | **0.094 ± 0.05** | **0.032** | **47.9** |
| 6 | experience\_yrs (年数) | 0.014 ± 0.01 | 0.009 | 29.5 |

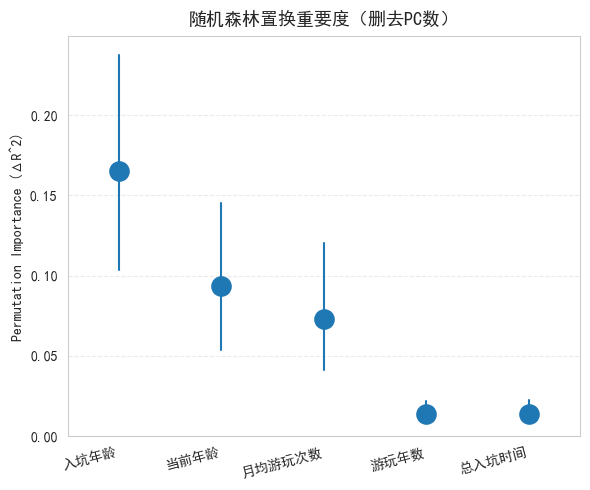
**本小节简单结论**

综合随机森林与 *XGBoost* 的多角度解释结果，可以归纳出以下关键发现。首先，正如预期，总游玩次数（play\_count）无论在置换重要度还是 *TreeSHAP* 贡献中都居于绝对核心地位，它构成了决定玩家最终Rating的“基础量”。

在 *XGBoost* 的 *TreeSHAP* 分解中，start\_age 平均能够左右约 118 分的评分变化，仅次于月均游玩次数（plays\_per\_month），在随机森林的 100 次自助抽样中，打乱 start\_age 令 R² 损失的区间稳定落在 0.10–0.24 之间，置信区间远离零点，显示其独立效应并非偶然噪声。



图表6 XGBoost (PI) 与XGBoost TreeSHAP

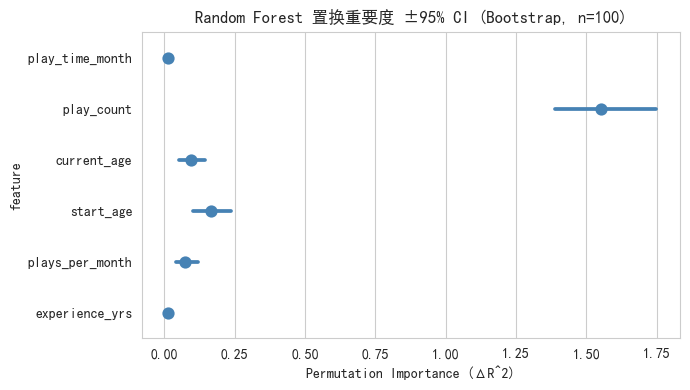
然而，在控制练习量之后，年龄相关变量依旧表现出清晰而稳定的影响，其中**尤以起步年龄（start\_age）最为突出**。换言之，越早开始系统练习的玩家，**其成绩“上限”显著更高**。即便两位玩家投入了相同的游玩次数，早入坑者依然平均高出数十至上百分。

其次，**当前年龄（current\_age）**的影响虽然弱于起步年龄，但仍然稳健可见。*TreeSHAP* 给出的平均贡献约 48 分，随机森林置换重要度的 95 % 置信区间落在 0.05–0.15 之间，说明年龄带来的生理与认知差异仍然会在高强度操作型节奏游戏中体现出来，只是幅度明显小于“早起步”所累积的经验红利。

图表 7（100次）随机森林结果（删除PC数）

此外，两种模型均提示**练习“密度”**优先于“时长”：在 *XGBoost* 中，月均游玩次数（plays\_per\_month）的贡献（≈ 167 分）和置换跌幅（≈ 0.13）都高于单纯累计月数或年数。

综合这些结果，可得一条清晰脉络：**大量练习奠定玩家下限，而越早入坑带来更陡峭的学习曲线与更高潜在上限；**当前年龄对表现的抑制虽存在，但不足以抵消早起步和高投入的优势。



图表8 （原始）100次Bootstrap 随机森林(置换重要度) ±95% CI

# 六、结论与建议

本研究以SEGA旗下《舞萌DX》街机音乐游戏玩家为样本，通过回收347份有效问卷，并结合实测游戏数据，系统分析了“年龄变量”（当前年龄、起步年龄）与“练习量指标”（累计游玩次数、月数等）对高水平游戏表现（DX rating）的影响。在统计建模和机器学习评估的基础上，我们得出以下主要结论：

首先，描述性分析与相关性热图显示，累计游玩次数（play\_count）和练习时长是决定DX rating的最主要因素，但玩家的入坑年龄（start\_age）与当前年龄（current\_age）也呈现显著的负相关，即越早接触游戏、年龄越小的玩家，其评分水平整体更高。

多元线性回归结果进一步证明，在控制了练习量的情况下，无论是当前年龄还是起步年龄，其主效应均为负，且均达到统计显著（m1模型，ΔR² = 0.021, *p* < 0.01）。而在引入“起步年龄 × 练习量”交互项后（m2模型），发现起步年龄对高强度练习者的表现影响更大，即越早入门且练习越多，评分上限越高。这一交互效应在后续分组可视化与局部加权平滑曲线（LOESS）中也得到了直观体现。

机器学习部分，随机森林与XGBoost模型对特征重要性的评估（包括置换重要度和TreeSHAP）一致表明：**累计游玩次数**始终是解释rating差异的最核心变量，其次为“起步年龄”和“当前年龄”。虽然年龄变量的解释度低于练习量，但在多种重要性评价指标下都排名靠前，表明其独立贡献不可忽视。分层可视化还揭示，部分14–16岁和19岁入门的玩家存在表现“回弹”，可能受限于青春期体力、专注力高峰以及大学生自由时间增加等因素。

综合来看，本研究验证了音乐游戏领域中的“天赋+努力”并存假说——**练习量是推动高水平表现的首要驱动力，但起步年龄与当前年龄在控制练习量后仍显著影响玩家的表现上限**。这一结论对青少年训练建议、赛事分级、以及音游社群运营均具一定参考价值。未来如能结合纵向数据或因果推断方法，将有望进一步厘清年龄效应的机制和边界。

# 参考文献

1. Bartholow B. et al. (2014). Reaction Time and Skill Acquisition in Youth Sports. J. Sports Sci.
2. Dodonova Y. (2022). Age-Related Performance Curves in Professional Esports. Frontiers in Psychology.
3. Watanabe D., Savion-Lemieux T., Penhune V. B. (2007). The effect of early musical training on bimanual coordination and timing. Brain and Cognition, 64(2), 152–163.
4. Bailey J. A., Penhune V. B. (2013). A sensitive period for musical training: Contributions of age of onset and cognitive abilities. Annals of the New York Academy of Sciences, 1252, 163–170.
5. Thompson J. J., Blair M. R., Henrey A. J. (2014). Over the hill at 24: Persistent age-related cognitive-motor decline in reaction times in an action video game. PLOS ONE, 9(4), e94215.
6. Ericsson K. A., Krampe R. T., Tesch-Römer C. (1993). The role of deliberate practice in the acquisition of expert performance. Psychological Review, 100(3), 363–406.
7. Macnamara B. N., Hambrick D. Z., Oswald F. L. (2014). Deliberate practice and performance in music, games, sports, education, and professions: A meta-analysis. Psychological Science, 25(8), 1608–1618.
8. Güllich A. (2017). International medallists’ and non-medallists’ developmental sport activities – a matched-pairs analysis. Journal of Sports Sciences, 35(23), 2281–2288.
9. Vaeyens R., Lenoir M., Williams A. M., Philippaerts R. M. (2009). Talent identification and development programmes in sport: Current models and future directions. Sports Medicine, 39(9), 703–714.....

# 附录 A：数据处理与分析代码概览

**1. 数据读取与特征工程**

1. **数据导入**：读取 dataset\_clean.csv，将 submit\_datatime 与 player\_birthday 转为 datetime 对象。
2. **年龄衍生变量**
   * current\_age = 报送日期 − 出生日期
   * experience\_yrs = 总游玩月数 ÷ 12
   * start\_age = 当前年龄 − 经验年数
3. **练习强度指标**
   * plays\_per\_month = 总游玩次数 ÷ 游玩月数
   * plays\_per\_year = 总游玩次数 ÷ 经验年数
4. **标准化评分**：计算玩家评分 player\_rating 的 *z-score* 存入 rating\_z。

**2. 描述性可视化**

1. **相关矩阵热图**：对 6 个核心指标绘制上三角相关系数热图，快速识别潜在共线关系。
2. **Lowess 平滑散点**：在 12–24 岁区间绘制 start\_age vs player\_rating，并以 LOESS 曲线揭示非线性趋势。
3. **分段着色 + 线性拟合**：按 play\_count 分箱上色，叠加 OLS 回归线，直观比较不同练习量下的年龄-评分关系。
4. **指数饱和模型**
   * 拟合 描述 *学习饱和*。
   * 再以对数横轴复现，验证指数转对数后呈近线性。

**3. 传统回归分析**

|  |  |  |
| --- | --- | --- |
| **模型** | **解释变量** | **目的** |
| **m0** | play\_time\_month + play\_count | 建立基准 |
| **m1** | m0 + current\_age + start\_age | 检验年龄主效应 |
| **m2** | m1 + start\_age:play\_count | 检验年龄×练习量交互 |

各模型均使用 statsmodels OLS；通过 F-test 比较 ΔR2ΔR^2ΔR2 评估增量解释力。

**4. 机器学习建模与特征解释**

|  |  |  |
| --- | --- | --- |
| **步骤** | **方法** | **关键配置 & 解释方法** |
| **4.1** | Random Forest | *Pipeline*：StandardScaler + 500 棵 RF；置换重要度（10 次） |
| **4.2** | XGBoost | *Pipeline*：缩放 + 参数化 XGB；同样计算置换重要度 |
| **4.3** | TreeSHAP | 直接调用 Booster.predict(pred\_contribs=True) 获取全局 SHAP 贡献 |
| **4.4** | 结果整合 | 双栏条形图同时展示置换 ΔR² 与 |
| **4.5** | Bootstrap RF | 100 次自助采样＋置换重要度生成 95 % 置信区间，为随机森林解释力提供稳健性度量 |

**5. 其他统计探索**

1. **主观问卷分组**：以箱线图比较玩家对“天赋重要性”(Question 1) 三档评价下的评分分布。
2. **K-Means 聚类**
   * 变量：start\_age、experience\_yrs、play\_count
   * 通过 WCSS 肘部法与 Silhouette 系数共同选择最佳 *k*；结果用年龄-评分散点上色呈现。
3. **分位数回归 (τ=0.1, 0.5, 0.9)**：考察练习量与年龄变量对不同评级分布尾部的影响差异。

**6. 评价指标速览**

* **解释力**：OLS R^2、RF OOB R²、XGB 置换 ΔR²
* **特征重要度**：Permutation Importance、TreeSHAP、Bootstrap CI
* **聚类质度**：Silhouette Score
* **显著性检验**：OLS t/F-test、分位数回归 *p-value*

**文件依赖**：

Python 3.10（miniconda），主要库版本 *pandas 2.x / numpy 1.26 / seaborn 0.13 / matplotlib 3.8 / statsmodels 0.14 / scikit-learn 1.4 / xgboost 2.0*。代码兼容 Mac/Win/Linux，图形渲染需中文字体支持。