

# 基于秩统计方法平衡科技人才项目评审 组间差异的实证分析

刘婧一

(北京干细胞与再生医学研究院, 北京 100101)

**摘要:** 通过梳理中国地方科技人才引进计划评审情况, 归纳和总结了科技人才项目评审的一般做法, 发现以分组同行评议为主流决策方式的人才项目评审面临不同组间评审主体差异的问题。基于此, 提出了一种评前预防和评后修正相结合的解决方法。评审前通过评价尺度、专家权重等方法对专家差异进行事先控制, 评审后通过正态秩分数法对评审主体差异进行均衡化处理, 从而有效平衡了组内和组间的评审主体差异。随机模拟验证与实证案例表明, 该方法能够在一定程度上平衡组间评审主体差异, 有利于提高同行评议决策的科学性, 对后续地方科技人才引进计划的立项评审工作具有一定的参考意义。

**关键词:** 人才项目评审; 评审主体差异; 秩方法

**中图分类号:** C931.2 **文献标志码:** A **文章编号:** 1671-1807(2024)22-0130-07

当今时代, 科技创新已成为引领和驱动经济社会发展的主引擎, 在现代化国家建设中具有核心地位。人才是创新驱动发展的根本所在, 增强国家创新实力离不开高端科技人才的支撑。党的十九届四中全会上提出“聚天下英才而用之”<sup>[1]</sup>, 要加快建设创新型人才队伍, 切实发挥人才在创新发展中的引领作用。党的二十大进一步强调, 要深入实施科教兴国战略、人才强国战略、创新驱动发展战略, 推动教育、科技、人才一体化发展<sup>[2]</sup>。随着国家对创新人才的定位越来越明确, 人才越来越成为推动经济社会发展的战略性资源, 为增强地区综合实力, 各地相继出台相关政策, 加大了对科技人才及其项目的引进和资助力度, 希望通过引进高层次人才和高质量项目, 推动科技创新和社会进步, 从而在国家经济发展中谋求一席之地。

鉴于科技创新人才对地方发展的重要性, 各地政府对引才计划立项评审的公平性、规范性和科学性十分重视, 在同行评议仍是评审主流形式的今天, 如何控制和平衡评审主体的差异, 提高评审结果的科学性, 对于评选出符合地方产业发展需求的人才具有重要意义。本文通过梳理中国地方科技人才引进计划的实施现状, 结合人才项目评审实践经验, 提出了基于秩统计方法解决不同组间评审专

家评分误差的方法, 为提高人才项目评审的科学性提供可行之法。

## 1 文献综述

近年来, 随着科技事业在国家发展中的战略地位不断提高, 越来越多的学者开始关注和探讨科技计划项目立项评审。其中, 同行评议这一主流方法也引发了众多探讨, 但研究大多从评审专家遴选与评价方面开展分析, 曹滔宇等<sup>[3]</sup>提出了一种基于贪心算法的科技项目评审专家多重匹配模型, 通过智能化手段从专家库中找出合适的评审专家组合团体, 规避人工遴选的纰漏。汪建等<sup>[4]</sup>基于元评价理论和实证数据建立了元评价综合模型, 分析专家评分质量, 从而指导专家库的动态调整。宋欢等<sup>[5]</sup>探讨了利用 ChatGPT 等 AI 大模型作为智能监测系统参与科技计划评审的可能性。金山等<sup>[6]</sup>提出了对评审专家的失信行为实时监控的智能系统解决方案。李彬等<sup>[7]</sup>从评审专家的专业水平、信用状况、评审态度以及承担科技计划项目能力四个维度设计了一套智能项目评审专家评价模型。杨健安等<sup>[8]</sup>基于相对逆序数的概念提出了排序导向科技评价中专家个体可信度的量化模型。张尧和林春<sup>[9]</sup>通过 Pearson 相似性分析方法测算专家项目评分一致性, 从而分析专家的可信度。

**收稿日期:** 2024-06-26

**作者简介:** 刘婧一(1994—), 女, 安徽利辛人, 硕士, 经济师, 研究方向为科技人才、科技管理和科技政策。

也有学者围绕专家评分尺度导致的误差开展研究,李樵和王菲菲<sup>[10]</sup>以双重加工理论为基础,探讨了评审人在创新性研究评价中的认知机制偏差,分析了认知偏差产生的原因,并从技术、管理、制度三方面,构建了干预认知偏差的有效策略。张金煊<sup>[11]</sup>提出了一种专家评价异常数据的修正模型及算法,用以识别异常数据,提升评审结果的有效性。李艳哲等<sup>[12]</sup>设计了一套固定评价要素和动态权重相结合的方法,对项目的评分赋予动态权重,以提升评审的客观性和评审环节的继承性。

对于平衡或消除组内评审主体差异的理论方法,现有研究大致分为两类:一是通过对数据进行标准化处理来消除主观评分误差,如首先利用统计模型,将评分数据转化成为具有一致性的数据,求得处理后数据的离散程度,并通过其分布确定临界值。若大于临界值则该数据不合理,可进行剔除或重新评价。二是在群组决策中对专家评分赋予权重。在大型群体决策的评价系统中,专家的知识水平、判断能力以及主观偏好的不同可能会导致评分出现系统误差,在群体决策中,如何客观确定多位专家的权重是关系到决策效果好坏的核心因素<sup>[13]</sup>。该方法主要遵循的原则是在不改变评价对象在评委心目中的地位的前提下,通过权重使得各专家的评分尽可能接近以减小评分系统误差及争议,提高群体共识度。群组决策中,修正专家评分的方法有主观赋权法、客观赋权法和主客观综合赋权法等<sup>[14]</sup>。

通过梳理相关文献发现,现有研究多聚焦专家评价的客观性、可信性和一致性,以及专家评分的组内差异,较少从实践的视角对平衡组间评分误差的方法进行研究。本文认为,对于组内差异来说,由于同组项目的评审主体一致,对于每个项目,即使仅是剔除极值后取专家分数的平均值,相对也具有一定可比性。在涵盖多领域多类别的大型评审活动中,相比组内专家差异,由评审主体不同而导致的组间差异更为突出。

本文主要探讨采取同行专家分组评议和排序决策法的地方科技人才项目评审中平衡组间评审主体差异的方法。基于当前科技人才项目评审实践,结合秩统计方法,探索出在规模性(即具有一定的统计学大样本意义)评审活动中可行性较高的一种方法,从而在一定程度上控制评审主体差异,提升评审结果的可信度和有效性,有助于政府和管理部门更加科学地利用评审结果选拔优秀人才。

## 2 “十四五”时期中国地方人才项目评审情况

### 2.1 政策导向

“十四五”期间,一线、新一线城市纷纷针对战略性新兴产业、未来产业、高新技术产业、科技金融服务产业等重点或特色领域,出台相应的科技人才引进计划,着力引进高层次创新创业人才、团队和项目。这些计划多以5年为一阶段,与国家的“五年计划”相契合。地方人才引进计划中,人才、项目、企业形成了“一体三位”的引进链条,政府通过实施人才引进计划,吸引高层次人才及其项目在本地落地和创办企业,从而提高区域创新能力,推动技术产业集聚,实现以科学技术驱动经济社会发展。

主要引才对象为海内外高水平创新创业人才,申报类别一般按照人才层次,如诺奖人才、领军人才、青年人才等,或按照科研活动性质,如创新人才、创业人才、科技服务人才等进行分类。根据不同的引才类型,可采用申报评审制、专家推荐制、政府邀约制等实施机制。申报周期一般以年度或滚动为主,入选者可获得几十万至几百万元不等的经费支持,部分地区对领军人才项目的支持力度达千万级。为确保引才计划效果,政府部门投入了大量人力、物力和财力,完善人才引进的配套生态环境,打造项目落地的配套产业链条。为保障各项经费投入发挥最大效力,切实引进符合地方发展需求的人才,立项评审环节至关重要。各地政府也更加关注评审方法的创新,以求最大程度保障引才计划的公平性和科学性。

### 2.2 主流评审方法

基于引才政策和相关实践,目前各地方在人才项目评审实践中多以群组评价为主要方法,通过讨论推荐、投票表决、评分排序等形式进行。地方政府和管理部门开展引才计划的主要目的在于为地方选拔符合地方发展要求的高层次人才,评审程序分一到多轮不等,两轮程序较为常见。其中,初评可通过采取资料审查、背景调查等形式进行;终评多采用同行评议法,通过函评、会议评审、调研评审等方式进行。

评价指标的设立应立足于引才计划的目标,一般从人才创新能力、项目技术水平、项目成果转化潜力、预期社会经济效益等方面设立一级指标,根据实际需求设立细分二级指标,并赋予每项指标不同权重。对于不同层次类别的人才,应设立各有侧重的评价指标进行分类评价。大部分地方引才计

划以引进创新创业人才为主,评审的主要关注点在于人才和项目的发展潜力和预期收益,与高校和科研院所侧重于引进基础性研究人才的计划有所不同。

评审实践中,由于申报数量较多,而且属于不同的人才类别和行业领域,所以首先根据一定的标准对申报人才及其项目进行分组,划分依据包括人才类别、行业领域等。根据不同组别的特征和需求,组建对应的评审专家组。评审中,每组专家按照事先设计的评审指标及标准,对人才项目进行打分,并给出定性意见。评审结束后,根据事先制定的规则对群组评价结果进行处理和筛选,确定最终结果。

对于运用评审结果筛选人才项目,排序法是较为主流的筛选方法,实践中有两种操作模式,以按照行业领域分组为例,一是分行业领域进行排序,确定入选名额后,按每个行业领域项目数量占项目总数的比例来确定每个行业领域的入选数量;二是不分行业领域对所有项目进行统一排序,划定一定的分数标准确定入选的人才项目。

为了避免行业争议,支持和平衡各类产业发展,地方引才计划的管理部门倾向于采取第二种方式,即不分行业领域对所有项目进行统一排序。在此种情况下,排名结果存在一些潜在的偏差。首先,不同行业对项目成熟度和发展潜力的认定有所区别,存在某一行业的项目分数总体偏高或偏低的可能。其次,对于同一组项目,由于评审主体评分的相对标准一致,故评分具有一定的可比性。但在不同组间,由于每组的评审主体群不同,在对所有人才项目的分数进行排序和比较时,可能存在由不同组间评审主体差异引起的误差。在此种情况下,对所有人才项目求取算术平均分进行统一排序可能会导致决策结果存在误差<sup>[15]</sup>。为了增强评审的科学性和分数的可比性,应采取一定的方法控制不同组间评审主体的差异。

### 3 平衡评审主体组间差异的方法

#### 3.1 评前预防——评价尺度法和专家权重法

在实践中,为提高评审结果的公平公正性,可采取建立统一标准并培训、确定专家构成及权重等方法对评审主体组间差异进行评前预防。

一是建立统一的评价标准并对专家进行专业培训。制定客观、清晰的评审准则、指标和标准,并通过评前培训提高专家对评审标准和方法的理解和掌握程度,尽量减少因理解差异导致的评分偏

差,从而确保所有评审主体都能在同一框架下进行评价,减少个人主观性评分差异。例如,根据引才计划的特点、评审需求,以及指标体系,设计若干分数等级区间,并设定每个区间的定性档次(如优秀、良好、合格、不合格)及详细的定性或定量解释。专家评审时,可以将定性档次作为评分的参考依据,一定程度上增强了不同组间评分范围的可控性和可比性。

二是设定评审组的专家构成及权重。在专家构成结构方面,按照领域、科研活动性质、专业水平、评审能力等因素,对数据库中的专家根据类别进行标记(表1)。在评审前,根据不同的人才项目评审需求,确定评审组的专家构成结构,如某人才计划旨在引进高层次创业人才,可提高具有创业和转化经验或科技投资领域的专家占比。

表1 专家分类案例

分类标准	类别维度
行业领域	前沿战略科技领域,或投资、财务、知识产权、成果转化等科技服务领域
科研活动领域	基础研究、应用研究、试验发展等
专业水平	代表性成果、职称、头衔、专业从事年限等
评审能力	一致性、稳定性、质量、态度等

在设定专家权重时,国内多采用专家的专业能力、评审经验、专家信誉、评审水平等作为指标为专家赋权<sup>[16]</sup>,通过权重对不同专家产生的系统误差进行综合评控。在统计分数时,对不同类别专家的评分赋予不同的权重,以区分其重要性。通过对专家进行分类和设置每组专家的构成结构,使得每组专家的认知标准和能力水平尽可能相近,从而在一定程度上控制评审主体差异。

#### 3.2 评后修正——秩分数法

##### 3.2.1 理论基础与研究设计

秩统计方法是一种基于排序数据的统计技术,秩即数据按大小进行排列的顺序,该方法的主要特点是利用数据的秩,而非实际数值来进行统计分析。由于秩方法构造的检验统计量在原假设下往往是与分布无关的,故尤其适用于不满足常规参数统计假设的数据情况,如非正态分布等。在科技人才项目评审情境下,考虑到“分组评价+统一排序”的现实需求,从秩统计方法的视角提出解决方案。

设某人才项目评审共有 $G$ 个评审组,每个评审组有 $m_i$ 位评委,且一位评委只属于一个评审组,每个评审组为 $x_i$ 个项目评分,且一个项目只属于一个评审组。设第 $i$ 评审组的第 $j$ 评委为此组的第 $k$ 个

项目评分为  $y_{ijk} (i = 1, 2, \dots, G; j = 1, 2, \dots, m_i; k = 1, 2, \dots, x_i)$ 。则共有  $x (x = \sum_{i=1}^G x_i)$  个项目,  $Y (Y = \sum_{i=1}^G m_i x_i)$  个专家评分。

$$G_i = \begin{bmatrix} y_{i11} & y_{i12} & \cdots & y_{ij1} \\ y_{i21} & y_{i22} & \cdots & y_{ij2} \\ \vdots & \vdots & & \vdots \\ y_{ik1} & y_{ik2} & \cdots & y_{ijk} \end{bmatrix} \quad (1)$$

设每个参赛队的平均分为  $\bar{y}_k = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ijk} (i = 1, 2, \dots, G; k = 1, 2, \dots, x_i)$ 。平均分最大值为  $y_{\max}$ , 最小值为  $y_{\min}$ 。

为去除不同评审组评分差异的影响,使得不同组别人才项目的评分具有可比性,引入秩分数的概念。

设有分数  $y_1, y_2, \dots, y_n$ 。从小到大排序后得到次序统计量  $y_{(1)} \leq y_{(2)} \leq \dots \leq y_{(n)}$ 。设  $y_m$  排在第  $d_m$  位 ( $1 \leq d_m \leq n$ ), 则令  $y_m$  的秩分数为  $d_m$ 。如果有  $L$  个  $y_m$  的数值相等, 根据以上规则计算的秩分数为  $d_m, d_m + 1, \dots, d_m + L - 1$ , 则取这些秩分数的平均值作为相同取值的  $y_m$  的秩分数, 即  $d_m + \frac{L-1}{2}$ 。

这样得到的秩分数取值区间为  $[1, n]$ , 与观测值个数  $n$  有关。为了  $n$  变化时的可比性, 把  $\{d_m, i = 1, 2, \dots, n\}$  变换到  $[0, 1]$ :

$$r_m = \frac{d_m}{n} \quad (2)$$

为了得到近似正态分布的秩分数  $z_m$ , 令

$$z_m = \varphi^{-1} \frac{d_m - 3/8}{n + 1/4} \quad (3)$$

基于此, 可将  $[0, 1]$  的秩分数  $r_m$  转换为正态秩分数  $z_m$ :

$$z_m = \varphi^{-1} \cdot \frac{r_m - \frac{3/8}{n}}{1 + \frac{1/4}{n}} \quad (4)$$

正态秩分数是比较接近于正常的分数分布的一种分布, 即分数处于中等水平的占大部分, 优秀和不及格的都占少数, 而  $[0, 1]$  内的秩分数更接近于分数排名。

基于以上秩分数的概念, 在专家评分结束后, 不分评审组, 把所有的  $Y$  个分数变换为正态秩分数  $Z$ , 然后对每个评委的分数进行调整, 使得每个评委的平均分等于其所在评审组的平均分, 再对每个评审组内的分数进行调整, 使得每个评审组的平均分

数与总平均分数一致。

最后计算调整后每个人才项目的平均分数  $\bar{z}_k$ , 平均分最大值为  $z_{\max}$ , 最小值为  $z_{\min}$ 。把  $x_i$  个人才项目的平均分数  $\bar{z}_k$  线性变换到  $[y_{\min}, y_{\max}]$ , 得到每个项目的最终得分。

$$y_{\text{final}} = y_{\min} + \frac{y_{\max} - y_{\min}}{z_{\max} - z_{\min}} (\bar{z}_k - z_{\min}) \quad (5)$$

将所有项目的最终分数按从高到低的顺序进行排序, 得到所有人才项目的最终排名。

### 3.2.2 对照方案与模拟验证

为了验证以上方案, 提出另一种仅考虑评审组内评委差异的方案, 即对每个评审组分别进行处理: 在第  $i$  组内, 先对每位评委分别计算, 第  $j$  位评委的  $x_i$  个分数转换为  $[0, 1]$  的秩分数  $r_{ijk} (k = 1, 2, \dots, x_i)$ 。基于此, 在第  $i$  组得到  $m_i x_i$  个秩分数, 然后计算第  $i$

组内第  $k$  参赛队的平均秩分数  $\bar{r}_k = \frac{1}{m_i} \sum_{j=1}^{m_i} r_{ijk} (k = 1, 2, \dots, x_i)$ , 将其转换为正态秩分数, 再线性变换到  $[y_{\min}, y_{\max}]$ , 由此得到  $x$  个项目的最终分数。

对以上两种方案开展了随机模拟测试: 对第  $i$  组 ( $i = 1, 2, \dots, G$ ), 设第  $j$  评委 ( $j = 1, 2, \dots, m$ ), 对本组第  $k$  个项目的评分为  $y_{ijk}$ , 按  $N(0, 1)$  标准正态分布随机模拟生成分数, 然后进行归一化处理, 使其在  $[40, 90]$ , 记为  $C_{ik} (i = 1, 2, \dots, G; k = 1, 2, \dots, x_i)$ 。

设各评审组评分有固定的偏差  $\alpha_i$ , 且  $\sum_{i=1}^G \alpha_i = 0$ , 即  $\alpha_i$  服从对称的随机分布。模拟时  $\alpha_i$  在  $\{\pm 2, \pm 5, \pm 10\}$  随机等可能选取, 取到其中任意值的概率相等。

设各评委评分有固定偏差  $\beta_j$ , 且  $\sum_{j=1}^m \beta_j = 0$ , 即偏差  $\beta_j$  服从对称的随机分布。模拟时  $\beta_j$  在  $\{\pm 2, \pm 4, \pm 8\}$  随机等可能选取, 取到其中任意值的概率相等。

设评委对人才项目真实水平的评分, 除固定偏差外, 还有随机误差  $\epsilon_{ijk}$ , 用  $N(0, \sigma^2)$  进行模型, 即随机误差  $\epsilon_{ijk}$  服从零均值、同方差的正态分布, 模拟时取  $\sigma = 4$ 。

建立线性混合模型来模拟分数数据:

$$y_{ijk} = C_{ik} + \alpha_i + \beta_j + \epsilon_{ijk} \quad (6)$$

随机生成数据时, 若模拟分数超过 100 则将其调整为 100 分。用以上两种方案对  $\{y_{ijk}\}$  进行处理后得到  $x$  个分数  $N_{ik} (i = 1, 2, \dots, G; k = 1, 2, \dots, x_i)$ 。  $\{N_{ik}\}$  产生的较高名次应该尽可能与  $\{C_{ik}\}$  产生的较高名次一致。

对每一组随机数据,分别用两种方案计算修正分数并进行排名。通过两项指标计算统计得到的名次 $\{N_{ik}\}$ 与实际的真实分数 $\{C_{ik}\}$ 应得的名次的一致性:

(1)名次差异的根均方误差。设 $r_{ik}$ 是由 $C_{ik}$ 得到的秩(名次),而 $\hat{r}_{ik}$ 是根据方案统计调整后得到的秩(名次),计算名次的根均方误差为 RMSE,根均方误差越大则此方案效果越差。

$$RMSE = \sqrt{\sum_{i=1}^G \sum_{k=1}^{x_i} (r_{ik} - \hat{r}_{ik})^2} \quad (7)$$

(2)名次的秩相关系数。计算 $r_{ik}$ 和 $\hat{r}_{ik}$ 的相关系数。相关系数越接近于 1 说明此方案效果越好。

选取一定的组别数( $G=7$ )、评委数( $m_i=8$ )、项目数( $x_i=10$ ),随机生成 100 组数据进行模拟和对比分析,两种方法的模拟效果如表 2 所示,对比名次

表 2 两种方案效果对比

方案		最小值	下四分位数	中位数	平均值	上四分位数	最大值	标准差	上下四分位差
根均方误差	实验组	3.8	6.1	7.1	7.1	7.8	14.2	1.6	1.7
	对照组	5.0	6.8	7.9	7.9	8.7	13.3	1.4	1.9
秩相关系数	实验组	0.751	0.925	0.939	0.939	0.955	0.983	0.032	0.029
	对照组	0.782	0.907	0.924	0.924	0.943	0.969	0.029	0.037

研究和成果转化的专家组建评委会。

在处理评审结果时,对所有专家独立评分进行描述性统计分析和极值剔除后,采用秩分数法进行分数处理,得出每个项目的最终得分和排名。处理结果与专家组原始分数算术平均值的对比如表 4 所示,调整前后每组入选项目的数量分布如图 1 所示。

差异的根均方误差和秩相关系数,发现实验组的效果优于对照组方法。

在实践中,可利用 R、Python 等程序,以上述原理为基础,制作出处理评审分数的软件工具,从而提高此方法的使用便捷性。

#### 4 实证案例

某地方出台了一项以引进高层次创新人才为目标的人才引进计划。根据相关政策,该计划以高额创业资金资助的形式吸引战略科技人才赴当地创办企业,开展科技成果转化。该计划每年度组织一次公开申报和评审工作。某年,申报并通过初评者共有 38 人,预期终评入选率约 40%。根据行业领域,将人才项目分为 5 组,每组的项目数量和评委数量如表 3 所示。由于该计划侧重于引进能够进行成果转化落地的科技人才,在遴选专家时,由不同领域内从事应用

表 3 某地方人才计划申报评审情况

组别	项目数量	专家数量
G01-生物医药	9	5
G02-新一代信息技术	8	5
G03-节能环保	11	6
G04-现代农业	3	4
G05-高端装备制造	7	5

表 4 修正前后的分数及排名对比

原始算术平均分						秩方法调整后分数					
排名	项目号	分数	排名	项目号	分数	排名	项目号	分数	排名	项目号	分数
1	G0403	91.3	20	G0504	83.8	1	G0310	91.0	20	G0107	84.7
2	G0205	90.0	21	G0502	83.4	2	G0205	91.0	21	G0507	84.5
3	G0310	89.8	22	G0105	83.2	3	G0103	90.1	22	G0504	83.9
4	G0201	87.4	23	G0204	83.0	4	G0403	90.0	23	G0306	83.5
5	G0503	87.4	24	G0306	82.8	5	G0308	88.3	24	G0204	83.5
6	G0308	87.3	25	G0505	82.6	6	G0201	88.2	25	G0502	83.5
7	G0207	87.0	26	G0309	82.5	7	G0303	87.8	26	G0106	83.4
8	G0303	86.8	27	G0305	82.3	8	G0207	87.8	27	G0309	83.1
9	G0203	86.8	28	G0104	82.2	9	G0503	87.7	28	G0402	83.1
10	G0506	86.6	29	G0307	81.3	10	G0105	87.7	29	G0305	83.0
11	G0501	86.4	30	G0102	81.2	11	G0203	87.6	30	G0101	82.8
12	G0311	85.8	31	G0206	81.0	12	G0506	87.5	30	G0108	82.8
13	G0103	85.4	32	G0107	80.4	13	G0311	86.7	32	G0505	82.6
14	G0208	85.0	33	G0304	80.2	14	G0501	86.7	33	G0109	82.4
15	G0402	84.8	34	G0106	79.2	15	G0104	86.7	34	G0401	82.3
16	G0302	84.7	35	G0101	78.6	16	G0208	85.6	35	G0307	81.9
17	G0507	84.4	36	G0108	78.6	17	G0102	85.6	36	G0206	81.4
18	G0301	84.2	37	G0109	78.2	18	G0302	85.5	37	G0304	80.7
19	G0401	84.0	38	G0202	76.0	19	G0301	84.9	38	G0202	76.0

注:以上分数均四舍五入保留 1 位小数点,如保留后分数相同,按保留前的实际分数排序。

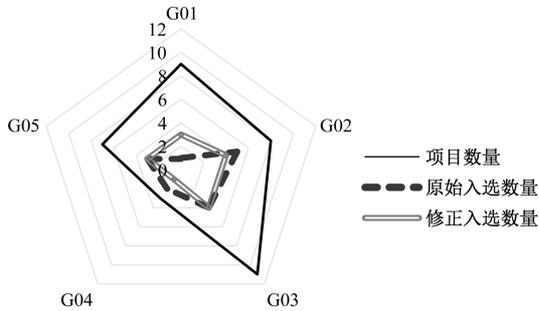


图1 分数修正前后的各组入选数量对比

此案例中,对于专家原始分数,第四组现代农业领域分数整体偏高,第一组生物医药项目分数整体偏低。秩分数法综合考虑了组内和组间差异,增强了不同组间分数的可比性,提升了评审结果的客观公正性。如表4所示,按照原始分数的平均值排名,由于第一组项目整体分数偏低,第一组最高分的项目排在第13位;经修正后,第一组最高分的项目排在第3位。由图1可以看出,修正后各组入选项目数量分布形态更加接近各组总体项目数量分布形态,由于个别组分数偏高或偏低而导致入选项目行业领域失衡的情况得以缓解。

为便于清晰地展示应用秩分数法调整前后的分数变化,本文选取仅有38个项目规模、组间项目和专家数量相差较大的案例进行举证。鉴于人才项目评审“分组评价+统一排序”的需求及秩方法的局限性,随机模拟验证和相关实践表明,为尽可能保证统计结果的科学性,此方法更加适用于人才项目总体规模在100及以上、最终入选比例高于25%、每组评审专家数量大于5的情形。

## 5 结论

在以同行评议为主的地方科技人才项目评审实践中,由于人才项目的行业领域和类别较多,通常会划分不同的组别,但在确定入选结果时,管理部门倾向于采取对所有组别的人才项目进行统一排序的筛选方式。由于组别间评审主体的认知标准不尽一致,导致不同组间的评分差异较为显著。而在学术研究中,消除组间误差差异的方法过于理论化和复杂化,难以应用到实践中。基于地方科技人才项目评审现状和“分组评价+统一排序”的需求,本文提出了一种可操作性较高的解决方案,即通过评价尺度、专家权重、秩分数等方法,对评审主体组内和组间差异进行综合平衡,使得评审结果更

加科学有效。

本文所提出的方法主要适用于采取同行专家分组评议和排序决策法的人才项目评审活动。在实践中,应根据不同引才计划的性质和目的,以及人才项目的数量等因素,采用不同的评审程序和方法,本文所提出的方法虽简便有效,但不一定适用于所有情况,实践中平衡和控制组间评审主体差异的普适性方法仍有待进一步探讨。

## 参考文献

- [1] 中国共产党第十九届中央委员会第四次全体会议公报[EB/OL]. (2019-10-31)[2024-01-02]. <https://www.12371.cn/2019/10/31/ARTI1572515554956816.shtml>.
- [2] 赵凌云. 加快建设教育强国、科技强国、人才强国[EB/OL]. (2022-12-08)[2024-01-02]. <http://cpc.people.com.cn/n1/2022/12/08/c448544-32582926.html>.
- [3] 曹滔宇,熊永平,史梦洁,等. 基于贪心搜索的分组项目评审专家遴选方法[J]. 计算机系统应用, 2021, 30(4): 168-174.
- [4] 汪建,王裴裴,丁俊. 科技项目专家评审的元评价综合模型研究[J]. 科研管理, 2020, 41(2): 183-192.
- [5] 宋欢,陈金辉,吴瑞林. 类ChatGPT模型介入国家科技计划项目评审的应用探讨[J]. 科技和产业, 2024, 24(5): 177-186.
- [6] 金山,邵之一,陈晓珊,等. 科技项目评审监测及行为智能分析研究与应用[J]. 计算机时代, 2021(5): 13-15.
- [7] 李彬,李旭红,应娟. 科技计划项目评审专家精准评价模型研究——以江苏省科技计划项目评审为例[J]. 南京师大学报(自然科学版), 2024, 47(1): 133-141.
- [8] 杨健安,刘传斌,余乐安. 基于相对逆序数的评审专家可信度评价方法[J]. 统计与决策, 2022, 38(3): 184-188.
- [9] 张尧,林春. 专家评审特征分析及专家反评估方法研究[J]. 工业工程, 2020, 23(6): 124-130.
- [10] 李樵,王菲菲. 认知视角下创新性研究的同行评议: 偏差与干预[J]. 科学学研究, 2021, 39(8): 1354-1363.
- [11] 张金焕. 科技评价中专家评审粗差数据的处理及仿真[J]. 计算机仿真, 2013, 30(7): 214-217.
- [12] 李艳哲,毛寒松,李珊. 基于多权值评价的科研项目评审方法研究[J]. 电子质量, 2021(7): 86-89.
- [13] 周伟,何建敏,余德建. 直觉模糊群决策中专家权重确定的一种精确方法[J]. 控制与决策, 2013, 28(5): 716-720.
- [14] 郭永辉,尚战伟,邹俊国,等. 群决策关键问题研究综述[J]. 统计与决策, 2016(24): 63-67.
- [15] 郭东威,丁根宏,毛俊诚,等. 确定专家权重的数量积法及在排名中的应用[J]. 烟台大学学报(自然科学与工程版), 2015(4): 249-254.
- [16] 胡雨欣,韩槌夏,陈状,等. 基于变权-理想点-潜在主题(VW-TOPSIS-LDA)的科技项目评审专家动态综合评价方法研究[J]. 科技促进发展, 2023, 19(12): 765-775.

## Empirical Analysis of Equilibrating Group Error of Assessment Subject in Local R&D Talent Project Review Based on Rank Statistics Method

LIU Jingyi

(Beijing Institute for Stem Cell and Regenerative Medicine, Beijing 100101, China)

**Abstract:** By analyzing the current situation of local R&D talent projects in China, the general practices of local R&D talent project review were summarized, and the problem of differences in the assessment subject between groups that current peer review activities face was revealed. In order to equilibrating group error of assessment subject, a combined solution containing pre-review prevention and post-review correction was proposed. Before the review process, methods such as setting evaluation criterion and expert weights could be applied. After the review process, rank statistics method could be used to convert original scores into normally distributed rank scores for equalization, and then linearly transform rank scores back to the original interval distribution, which would effectively balance the differences in assessment subjects within and between groups. Random simulation and empirical practices show that this method can balance the differences of evaluation subjects between groups, which will help improving the objectivity of the score processing and providing certain reference to future local talent project review.

**Keywords:** local talent project evaluation; differences of assessment subject; rank statistics