

关于杰出科学家科研高峰年龄的再探索

沈楠¹,汪琛²

(1. 安徽农业大学外国语学院,安徽合肥 230036;2. 中国科学院科技战略咨询研究院,北京 100190)

摘要:数据密集型研究范式为科学学研究开辟了新路径,使得诸多经典命题得以在更宏大的实证维度上重新审视。本研究聚焦科学社会学中关于科研生命周期的年龄问题,深入探讨“科学家科研高峰年龄统计分布”这一传统论题的新结论,研究发现与赵红州等学者提出的“科学发现年龄定律是一种威布尔分布”不尽相同,基于传统数据集扩张后得出的最优拟合分布为“正态逆高斯分布”。进一步分析表明,数据集的最优拟合分布具有非静态特征,不同的数据采样标准会导致数据集在内涵与外延上存在差异,从而使得不同数据集可能对应不同的最优拟合分布。同时,本研究对不同年代数据集的均值年龄、中位数年龄、最小年龄、最大年龄等局部指标进行描述性分析和思考。在此基础上,本研究认为学术共同体应更加关注统计规律与微观个案之间的辩证关系,在重视一般性科研高峰年龄统计规律的同时,也要深入挖掘具体科学家个案中蕴含的丰富典型化意义。

关键词:杰出科学家;科研高峰年龄;统计规律;个案考察

中图分类号:N031

文献标识码:A

众所周知,纵观一部精彩纷呈的科学发展史,科学发现的过程似乎有其微妙的内在规律性。对于任何科研工作者而言,都需要经过不同程度的知识和经验积累后才能实现由量变到质变的飞跃。鉴于此,杰出科学家取得重大科技成果时的生理年龄和职业年龄的分布问题一直是学者们的关注焦点,而相关结论在大科学时代的当下,依然有其独特的参考价值 and 实践意义。事实上,回溯过往科学学领域的已有研究成果可以发现,科研人员在其人生各个阶段都有可能取得重大突破。例如,在备受关注的诺贝尔科学奖项获得者中,英国物理学家威廉·布拉格(William Bragg)年仅24岁就创立了X射线晶体结构分析方法,并在一年后荣膺1915年度的诺贝尔物理学奖,成为历史上最年轻的诺贝尔科学奖获得者(时年25岁);与之相对,电池材料专家约翰·古迪纳夫(John Goodenough)在75岁高龄时仍然“宝刀未老”,成功研发锂离子电池制造中的关键材料磷酸铁锂,并于2019年斩获诺贝尔化学奖,成为历史上最年长的诺贝尔科学奖获得者(时年97

岁)。

为了进一步探究科学家发明创造峰值年龄的最优统计分布,长期以来很多专家付出了艰苦卓绝的努力,也取得了不少学术界耳熟能详的精彩成果。就国内相关研究而言,赵红州等人开展了非常关键的初创工作,基于定量统计方法提出了人们熟知的“科学发现最佳年龄定律”“人类重大科技成果的威布尔(Weibull)分布”等经典结论^[1-4],在此基础上国内学者亦进行了一系列拓展研究^[5-9];除此以外,诸多国内外学者亦依据不同类型数据集引入了更为多样化的研究视角^[10-12],并且提出了关于科研产出高峰年龄的各类统计模型,除了国内学术界较为熟悉的“威布尔分布”“lognormal分布”之外,较有影响力的研究成果还有“文献生命周期对数正态分布模型”^[13]、“重大成果指示器的最小二乘回归方程”^[14]、“科研创造力在不同学科以及不同年代的年龄动态模型”^[15]、“两步认知模型”^[16]、“诺贝尔经济学奖获得者的科研创造峰值年龄变化模型”^[17]等等。

收稿日期:2024-09-11;修回日期:2025-06-12

基金项目:国家社会科学基金重点项目(23AZD040);中国博士后科学基金第75批面上资助(2024M751746);安徽省哲学社会科学规划项目(2025-674)

作者简介:沈楠(1980-),男,副教授,博士。汪琛(1996-),男,助理研究员,博士后,通讯作者,E-mail: chen0522@mail.ustc.edu.cn。

1 问题提出

值得注意的是,以上统计模型在提供重要论断的同时,也不可避免地存在一些不足,集中体现在科研产出衡量指标的选取标准上。例如,不少学者在开展数据采样的过程中,普遍关注“诺贝尔科学奖获得者”这一较为权威的科学精英群体,并以此为依据建立各类数据集。然而,尽管诺贝尔奖被视作当今全球最具影响力的科学奖项,其遴选范围却较为狭窄,仅覆盖了物理学、化学、生理学和医学等领域,未能纳入数学、地质学、临床医学、工程科学以及计算机科学等重要领域。此外,如果仅仅以是否荣膺重大科学奖项为统计指标,就难免会忽略那些曾做出重大发现或发明、却未能得到任何奖项青睐的杰出学者。这里有必要强调的是,这类“被湮没者”在人类科技史上绝非可有可无的罕见个案,而是值得学术界重视和去深入发掘的研究对象。

鉴于此,本文尝试在既往相关科学家年龄研究的基础上,引入数据发掘的动态思路,即在建立数据集的过程中不仅限于纳入那些曾经荣获诺贝尔奖级别的精英学者,同时广泛发掘源自各个学科、以各种不同方式被学术共同体予以认可的杰出科学家们。此外,考虑到现代计算机软件能够开展远为丰富且精确的统计分析,本研究自建数据集的相关处理软件由本文作者独立编写完成,软件编写语言为 python,其功能在于分析自建数据集中所有科研高峰年龄样本的最佳拟合统计分布,并且将其与赵红州、梁立明、刘俊婉等学者的研究结论进行比较,以期得到若干具有一定启发价值的校正性结论。值得一提的是,本软件在运行中经历了数轮自主学习训练,因此能够确保得到较为可靠的分析结果。

2 数据集建立

鉴于以上论述,笔者整理并统计了 680 位取得重大基础科学发现或技术发明成果的杰出科学家科研高峰年龄数据,与以往相关研究建立的科研高峰年龄数据集相比,本数据集具有以下特点:

(1)数据集以相关科学家取得杰出贡献时的生理年龄为样本,不仅样本数量较大,其涉及的时间序列跨度也较长,从文艺复兴时期的尼古拉·哥白尼(Mikołaj Kopernik)提出日心说直至 2022 年赛缪尔·

艾尔特曼(Samuel H. Altman)领导团队开发新一代人工智能模型 ChatGPT。此外,考虑到科研成果诞生与科研成果发表之间的时间差,笔者充分调研了大量史料,将数据样本尽量控制为科学家萌发重大科研成果的时间,从而减少数据集的统计偏差。

(2)数据集涵盖的学科门类更为综合全面,不仅参考了诺贝尔奖涉及的物理学、化学、生理学和医学领域数据,同时还以菲尔兹奖和图灵奖的相关获奖者信息为基础,全面考察近代以来最富影响力的数学以及计算机科学成果,并且尽可能纳入若干具有重要历史影响力、被教科书载入的其他学科成果以及技术发明数据,例如“创立大陆漂移说(地质学)”“测定地球年龄(地质学)”“创立现代信息论(信息科学)”“发明电子计算机(信息科学)”“发明心脏导管术(临床医学)”“发明工业机器人(工程科学)”等等;此外,诚如梁立明、武夷山等学者所指出的:“严格说来,文献[1]、[2]、[3]和本文采用的数据都是西方科学史数据。但是,我们又注意到,东西方国家在社会历史、文化传统、教育体制和科学家、艺术的思维方式等方面是有一定差别的。这种差别会影响东西方国家科学与文化成果整体的年龄分布规律吗?这是一个有趣的问题”^[18]。鉴于此,本数据集的建立亦高度重视采集华罗庚、吴文俊、陈省身、钱学森、钱三强、侯德榜等中国杰出科学精英的科研高峰年龄样本。

(3)此外,数据集建立亦高度重视那些未被诺贝尔奖认可的同领域杰出科学家信息,例如创立玻色-爱因斯坦统计理论的萨特延德拉·玻色(Satyendra Bose)、证明生物遗传物质是 DNA 的奥斯瓦尔德·艾弗里(Oswald Avery)、成功拍摄高质量 DNA 晶体衍射图片的罗莎琳·富兰克林(Rosalind Franklin)、首次发现脉冲星的剑桥大学研究生乔瑟琳·贝尔(Jocelyn Bell Burnell)、发现星系转速恒定性的薇拉·鲁宾(Vera Rubin)、研发脊髓灰质炎疫苗的乔纳斯·索尔克(Jonas Salk)和阿尔伯特·沙宾(Albert Sabin)、发现 B 淋巴细胞和 T 淋巴细胞的麦克斯·库伯(Max Cooper)与雅克·米勒(Jacques Miller)等等。

(4)同样出于数据完整性考虑,本数据集也尝试发掘若干曾因各种缘由被不幸埋没、后经科学史家重新发现的杰出学者信息,例如发现天文学中“勒维特定律”的亨丽艾塔·勒维特(Henrietta Leavitt)、和导师共同发现链霉素的美国罗格斯大学

研究生艾尔伯特·沙茨(Albert Schatz)、发明世界上第一台“阿塔纳索夫-贝瑞计算机”的约翰·阿塔纳索夫(John Atanasoff)和克利福德·贝瑞(Clifford Berry)、开创生物节律分子机制研究的罗纳德·科纳普卡(Ronald J. Konopka)、首次发现断裂基因的华裔科学家周芷(Louise T. Chow)、首次克隆绿色荧光蛋白基因的道格拉斯·普瑞舍(Douglas Prasher)、开发克隆羊技术的基思·坎贝尔(Keith Campbell)等等。

总而言之,通过对以上数据集的统计分析,本文希望能够进一步探析科学家发明创造峰值年龄的结构分布,从而更加精准地揭示人类科技创新活动中的潜在规律。

3 数据集处理

首先,计算机软件针对数据集的“数据集中趋势”以及“数据离散程度”展开分析,并得到以下结果:

表 1 相关数据的集中趋势分析表

Table 1 Central tendency analysis of relevant data					
变量名	样本量	最大值	最小值	均值	中位数
年龄	680	83	19	38.875	38

表 2 相关数据的离散趋势分析表

Table 2 Analysis of dispersion trends of relevant data				
变量名	样本量	方差	标准差	变异系数
年龄	680	98.339	9.917	0.255

表 3 拟合效果最佳的前 5 类分布

Table 3 The top 5 distributions with the best fitting effects				
分布	总误差平方和	BIC	Ks_static	Ks_pvalue
正态逆高斯分布	0.029618	-6802.11762	0.034785	0.376973
Nef 分布	0.029629	-6795.34125	0.034234	0.397146
α 分布	0.029630	-6808.35960	0.036234	0.327412
F 分布	0.029630	-6801.83420	0.034147	0.400411
反 Gamma 分布	0.029632	-6808.31733	0.034303	0.394599

在此有必要说明的是:“总误差平方和”即指拟合分布与样本数据之间的误差,其数值越小则说明拟合效果越理想;“BIC”为“贝叶斯信息准则”(Bayesian Information Criterion)的缩写形式,其数值越小则代表拟合效果越好;在 Kolmogorov-Smirnov

根据相关方差和标准差的数值,不难判断科学家发明创造峰值的年龄数据具备较高的离散程度。简言之,尽管数据平均值为 38.875,但是在这一均值附近浮动的样本数量仍然较为可观,从而印证了数据集的结构分布具有显著的多样化特征。此外,相关变异系数值超过通常临界值(0.15),达到 0.255,这一现象说明当前数据集中存在若干“异常值”。然而,该变异系数值不大,说明当前数据集中严重偏离平均值的样本数量不多,大多数都接近平均水平。图 1 为基于当前数据集绘制的直方图,借此能够直观考察所有数据的整体分布情况。

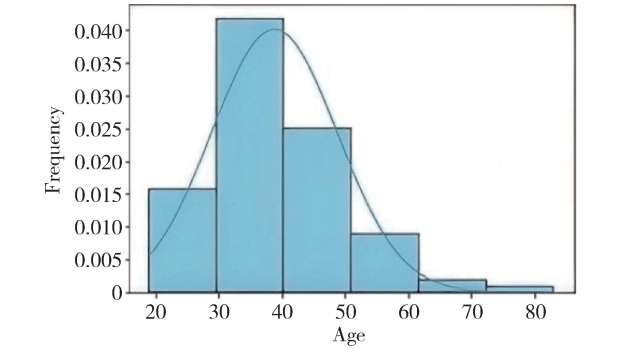


图 1 当前数据集直方图

Figure 1 Current dataset histogram

结合以上直方图性状,不难观察到数据整体分布呈现右偏态特征。为了进一步找到最优拟合该数据集的分布函数,计算机软件借助 Fitter 库遍历了当前收录的所有统计模型,并自动选取拟合效果最佳的前 5 类分布,其相关检验参数如表 3 所示:

检验中,Ks_statistic 表示 KS 统计量,其数值越小代表数据与理论分布之间的差异越小;Ks_pvalue 则表示 KS 检验的 p 值,其代表观察到的数据与理论分布之间的差异是由随机因素造成的概率。如果 p 值小于显著性水平(通常设定为 0.05),则拒绝原假

设,认为数据与理论分布之间存在显著性差异。此外,尤为值得关注的是,就当前数据集而言,根据软件处理结果,Weibull分布与当前数据集的拟合 p 值为 0.000,因此无法完成拟合任务。

综合以上分析,不难得出如果以“总误差平方和”为主要参考指标,则具备最佳拟合效果的为正态逆高斯分布 (norminvgauss),其分布图像如图 2 所示:

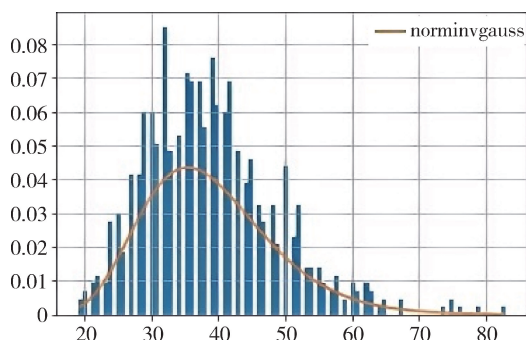


图2 正态逆高斯分布图

Figure 2 Norminvgauss distribution graph

在统计学中,正态逆高斯分布 (norminvgauss) 是一种连续性概率分布,其概率密度函数 (PDF) 为:

$$f(x; a, b, \delta, \mu) =$$

$$\frac{a}{2\delta\Gamma(1/b)} \exp\left[\frac{b(x - \mu)}{\delta}\right] \exp\left[-\frac{\exp\left(\frac{b(x - \mu)}{\delta}\right)}{2}\right]$$

其中, μ 和 δ 为分布的均值和标准差, a 和 b 为形状参数。结合当前数据集进一步拟合计算后,四个参数值分别为: $a = 107.048$, $b = 105.804$, $\mu = -0.302$, $\delta = 6.027$,完整形式的正态逆高斯分布函数可表示为:

$$f(x; a, b, \delta, \mu) =$$

$$\frac{107.048}{2\delta\Gamma(1/105.804)} \exp\left[\frac{b(x + 0.302)}{6.027}\right] \exp\left[-\frac{\exp\left(\frac{b(x + 0.302)}{6.027}\right)}{2}\right]$$

4 数据集最优拟合分布的非静态特征

首先,从当前数据集的统计分析结果来看,的确存在某个所谓的“科研高峰年龄最优分布区间”,其大致可以界定在“30 - 45 岁”之间,而这一推定也与长期以来的既定研究结论相吻合,即人类科技史上

的大多数重要研究成果都是科学家们在其青壮年时期取得的。

其次,正如前文所言,相比诸多前期研究而言,当前数据集的采样标准有了较大变化,其一方面扩充了学科门类,广泛纳入了数学、力学、地质学以及计算机科学领域的诸多成果,同时还在科学精英类型的多样化方面有所考量,不局限于以往仅被世界顶尖科学大奖所认可的科学家,而是将视域延拓至最近几个世纪以来所有在人类科技史上留下痕迹的杰出学者。对当前数据集进行分析后,笔者发现其最优拟合函数既不同于刘俊婉提出的对数正态分布(数据集为共计 435 位诺贝尔自然科学奖获得者的重大科学发现年龄),与赵红州等学者的观点——“科学发现年龄定律是一种威布尔分布(数据集为 1500 - 1960 年的共计 1928 项重大科学发现)”^[2]亦存在相当差距。因此,赵红州等人在 20 世纪 90 年代初提出的经典结论似乎已不能视为定论,而是存在进一步的校正和完善空间。

鉴于以上分析,本文认为不同的数据采样标准会导致数据集在内涵与外延上的若干差异。而对于“科学家的科研高峰年龄分布”这一科学学领域的经典论题而言,似乎不存在某个固定、静态化的单一结论。由于不同学者对于“杰出科学家”的界定时常缺乏统一共识,借此建立起的不同数据集则完全可能对应不同的最优拟合分布,而这一点在赵红州的相关论述中已经有所提示:“我们的工作验证,而不是证明,故而结论并不能作普遍的推广。若要真正证明(或否认)近、现代重大科技成果年龄分布是威布尔分布的规律,还必须从科学知识结晶学、科学能力学以及生理学和心理学等诸多方面进行研究”^[3]。

因此,尽管本文已经竭尽所能,同时采用网络数据挖掘以及人工识别方法采集相关信息,可是当前数据集在样本规模以及类型上的不足仍然在所难免。学者们在后续研究中如果设计出样本规模更大、样本类型更为丰富的数据集,则完全有可能探索出更加丰富和深刻的动态化结论。

5 数据集的若干局部特征分析

事实上,就当前数据集而言,所谓的“最优拟合分布”是其宏观性质,而其内蕴的若干局部特征也同样值得仔细甄别。由于当前数据集样本的时间跨

度较大,因此不妨选择不同年代的科学家创造峰值年龄数据开展进一步研究,选择均值年龄、中位数年龄、最小年龄、最大年龄进行描述性统计分析。

5.1 不同年代的均值年龄数据比较

从表 4 的数据呈现趋势来看:从 16 世纪到 20 世纪科学家取得重大突破的平均峰值年龄差异并不显著;然而,21 世纪的平均峰值年龄却高达 47.056,相比之前几个世纪的数据增幅明显。其原因在于迄今为止采集到的 21 世纪相关样本数量较为有限,且其中包括在引力波探测领域取得卓越成就的 3 位高龄科学家,分别为时年 83 岁的雷纳·韦斯(Rainer Weiss)、75 岁的基普·索恩(Kip Thorne)以及 79 岁的巴里·巴里什(Barry Barish),因此导致 21 世纪的数据平均值远高于之前年代水平。

表 4 不同年代的均值年龄数据

Table 4 Mean age data of different decades

年代	平均峰值年龄
1500 - 1599	40.000
1600 - 1699	42.357
1700 - 1799	41.471
1800 - 1899	38.606
1900 - 1999	38.987
2000 - 2022	47.056

5.2 不同年代的中位数年龄数据比较

借鉴同样方法,也可以针对不同年代科学家发明创造峰值年龄的中位数进行分析,以下即为不同年代的中位数数据。从数据趋势来看,从 16 世纪到 21 世纪的中位数年龄在整体上处于正常的数据浮动水平。

表 5 不同年代的中位数年龄数据

Table 5 Median age data of different decades

年代	中位数年龄
1500 - 1599	40.000
1600 - 1699	41.500
1700 - 1799	46.000
1800 - 1899	37.000
1900 - 1999	39.000
2000 - 2022	45.500

5.3 不同年代的最大科研峰值年龄数据比较

同理,亦可针对不同年代科学家发明创造峰值年龄的最大值进行分析,表 6 为不同年代的最大峰值数据。

表 6 不同年代的最大科研峰值年龄数据

Table 6 Maximum peak age data of scientific research output of different decades

年代	最大峰值年龄
1500 - 1599	51.000
1600 - 1699	61.000
1700 - 1799	54.000
1800 - 1899	76.000
1900 - 1999	75.000
2000 - 2022	75.000

从相关数据趋势来看,16 - 18 世纪的最大年龄数值水平偏低,而 19 - 21 世纪的数值水平增幅明显。有必要说明的是,导致以上现象的原因是多方面的,历史发展阶段、特定文化背景、社会教育水平、科研投入力度等因素都能够发挥不同程度的作用。然而,人类科学建制在 19 世纪之后的逐渐巩固与完善或许是根本原因。众所周知,早在 19 世纪之前,科学研究与技术发明在很大程度上还归属于个人兴趣爱好的范畴,缺少来自国家体制的有力支持和保障。有志于科技研发事业的业余爱好者们只能想方设法自行筹措所需资源,可以说整个过程相比于今天而言要困难和曲折的多,这种状况也导致当时人们的可持续性科研能力普遍无法与今日比拟。

受限于个人有限的时间、精力和财力,除非得到有产者的强力扶持,一个人的业余科研事业恐怕最多捱到 50 岁就不得不偃旗息鼓了。反观今日,由于可以得到源自国家的体制性收入保障,亦能以各种方式获取支撑科研工作的各项财务和组织资源,科学家的职业生涯可持续性也理所当然地延长了很多,在高龄阶段取得重大科研发现的概率也自然水涨船高。在今天,年逾 70 的资深科学家仍然像年轻学者那样终日忙碌在教科研事业中,这一现象在现代科研共同体中早已司空见惯。

5.4 不同年代的最小科研峰值年龄数据比较

最后,本文针对不同年代科学家发明创造峰值年龄的最小值进行分析,表 7 为不同年代的最小值数据。

从相关数据趋势来看,16 - 17 世纪以及 21 世

纪科学家取得重大突破的最小年龄为 30 岁左右,18-20 世纪科学家取得重大突破的最小年龄则仅为 20 岁左右。然而,有必要指出的是,这一数据水平很难被认为具有任何特定意蕴。尽管就统计意义而言,科学家的学术高峰年龄普遍分布在 30-50 岁之间,作为单独的个体每位科学家却都有自己的成长速度和职业发展之路。对于那些具备早慧和天才特质的学者个体来说,何时会迎来自己的科研爆发高峰期,在很大程度上是无法提前预知的随机化事件。此外,由于现代科学教育以及知识积累周期的大幅延长,专家们普遍认为自从 21 世纪之后,如同埃瓦里斯特·伽罗瓦(Évariste Galois)、沃纳·海森堡(Werner Heisenberg)那样年方弱冠就技惊四座的经典案例恐怕很难再出现了。

表 7 不同年代的最小科研峰值年龄数据
Table 7 Minimum peak age data of scientific research output of different decades

年代	最小峰值年龄
1500 - 1599	29.000
1600 - 1699	28.000
1700 - 1799	19.000
1800 - 1899	20.000
1900 - 1999	20.000
2000 - 2022	29.000

6 “统计规律”与“微观个案”

长期以来,所谓的“科学发现最佳年龄”对于我国的相关科技人才政策发挥了极为重要且微妙的影响,其直接表现就在于我国各级别的科研人才项目都明确了申报年龄限制,并将其设定为任何人都无法僭越、极为严格的硬性标准。事实上,该举措作为一种大科学时代的宏观控制手段,在很大程度上参考了学术界关于科研高峰年龄最优分布的研究结果,我国在科研精英遴选过程中也体现出日益显著的年轻化趋势。然而,考虑到每位科学家个体在创新过程中彰显出的丰富个性与异质化特征,在杰出人才的甄选上如果过于拘泥年龄方面的刚性限制,其实同样可能构成对科学家个体作用的强烈抑制,也不符合科技创新活动内蕴的本质规律。

众所周知,在当下我国科学精英的甄选和培育

机制中,已经形成了愈发显著的“线性进阶”模式(就读名校→受到名师提携→入选“四青”等国家级青年人才项目→成为“杰青”或“长江学者”→当选“两院”院士)。例如,以在我国最受瞩目的“杰青”项目和“院士”称号获得者为样本,笔者统计了从 2001 年至 2021 年间,我国历届新增中国科学院院士与“杰青”之间的身份耦合关系,从图 3 中可以明显观察到:大约自 2011 年以后,我国大半新增院士都主持过“杰青”项目,且所占比率有逐年升高趋势。从近几次的中国科学院院士增选情况来看,50 岁左右的中青年院士几乎被“杰青”们垄断,从而彰显出“杰青”和“院士”之间清晰可辨的衔接与同构进阶特征。

鉴于以上数据之间的耦合关系,似乎不难得出以下结论:在当前我国高级别科研人才的成长序列中,一个青年学者要在其职业生涯早期取得尽可能丰硕、具备足够显示度的成果,才有希望持续跻身于不同层级的精英身份竞争。在此情形下,诸多年轻学者们不得不陷入一轮又一轮残酷无情、淘汰率极高的学术锦标赛之中。因为高级别人才项目之间环环相扣的“闭环”特性,势必会造成“不入选即淘汰”“一步慢步步慢”的高强度竞争模式。此外,亦有学者统计了“杰青”与“四青”人才(“青年长江学者奖励计划”“优秀青年科学基金”“万人计划——青年拔尖人才支持计划”以及“青年千人计划”)之间的衔接关系,发现“当选年龄越低的‘杰青’群组,‘四青人才’的比例越大。在 40 岁以下的‘杰青’中,‘四青人才’的比例明显高于整体水平,从 2014 年开始已经超过一半,达到 53.19%,到 2017 年,这一比例已高达 95.24%。‘四青人才’成为‘杰青’重要的人才库。可以预期,随着‘四青人才’的持续推进及中选者的成长,‘杰青’中来自‘四青人才’的数量将会持续增长”^[19]。由此,按照科研精英的以上发展路径,科研工作者如果在 35 岁之前还尚未被纳入讲究环环相扣的体制化进阶轨道,那就意味着其很可能终其一生都失去了成为精英人才的有效竞争力。

有必要强调的是,此类现象虽然貌似严酷,却符合我国学术界关于科研最优年龄区间的既定认识以及政策设计者的原始意图,即尽早选拔出资质过人的青年才俊并给予重点倾斜式培养。可是,纵观一名科研从业者的整个职业生涯,那些最佳科研成果并非总是诞生于所谓的“最优年龄区间”。立足于

微观场域而言,也不是所有的精英科学家在其职业生涯早期都能快速崭露头角。例如:2004 年诺贝尔物理学奖得主弗兰克·维尔扎克(Frank A. Wilczek)发表的首篇论文就成为其摘取诺奖的主要依

据。与此相对应,2002 年诺贝尔化学奖得主约翰·芬恩(John B. Fenn)直至被耶鲁大学强制退休后才发表了相关诺奖论文。

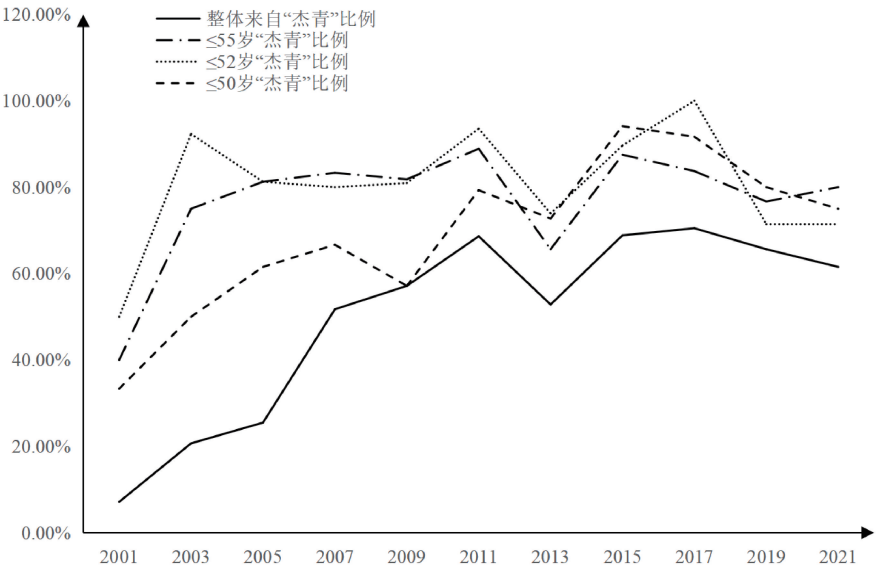


图 3 2001 - 2021 年中科院新增院士源自“杰青”比例表
Figure 3 Proportion of New CAS Academicians voted out of “Outstanding Young Scientists” from 2001 to 2021

事实上,前述现象在科学史上亦不罕见。艾尔伯特·巴拉巴西(Albert - László Barabási)等学者在综合分析横跨 7 个学科、多达几十万条论文出版信息的基础上,总结出了颇具启发意义的“个人成果影响力随机分布规律”(random - impact rule):“我们发现在一个科学家的职业生涯中,他发表的最高影响力作品是随机分布在其全部科研产出中,也就是说所谓的最高影响力作品会以同等概率分布在科学家发表的全部论文序列中的任意一个位置——它或许是科学家发表的第一篇论文,也可以出现在科学家职业生涯的中期,也可能是其发表的最后一篇论文。这个规律适用于任何科学家类型,无论其专业学科、职业生涯年限、所生活的年代、从事个人研究亦或加入某个团队……”^[20]。

综上所述,在数理统计学视阈下总结出的所谓“科学发现最佳年龄”,究竟在多大程度上能够发挥其本身应有的参考价值和意义?我们又该如何理解抽样统计研究与具体案例研究之间的关系呢?其实,与那种随机抛硬币再计算其落地正反面概率的统计实验不同,当前数据集尽管在整体上呈现出正态逆高斯分布趋势,数据集中的每个具体样本却都

与一位活生生、有血有肉的杰出科学家相关联,无论其展现出的科研高峰年龄大小如何,都曾在一部波澜壮阔的人类科学史中留下无法磨灭的印记,也由此具备了历史、文化与科学视域下的多重内涵,同样值得研究者的高度关注。

在科研道路上当然不乏“自古英雄出少年”的华美篇章,然而,那些一生追求真理、老骥伏枥、志在千里的“老龄理想类型”又是否真的可以缺少? 67 岁证明生物遗传物质是 DNA 而非蛋白质的奥斯瓦尔德·艾弗里(Oswald Avery); 65 岁提出深度学习算法、被誉为当代最伟大人工智能学者的杰弗里·辛顿(Geoffrey Hinton); 75 岁高龄研发出磷酸铁锂材料的锂电专家约翰·古迪纳夫; 身处重重逆境仍深藏功名、以 58 岁大龄讲师的身份在“孪生素数猜想”上取得破冰成就的华人数学家张益唐; 被耶鲁大学强制退休后仍坚持研究、67 岁发明“电喷雾离子化法”的约翰·芬恩……

鉴于此,纵使在科研共同体已经高度体制化的今天,借助那些细致的个案研究,我们依然能够发掘出若干与前述科学精英培养“年轻、早熟化模式”大相径庭的精彩案例。或许,这些案例不具备抽样统

计意义上的代表性,但在广阔的科学精英群体中却具有某种特殊的“非常规”典型性。就以上角度而言,本文自建数据集中的具体样本又都超越了单纯作为统计数字的狭隘意义,乃至被赋予某种内蕴丰富个性化因子的社会历史视域,而所谓“统计规律”与“微观个案”之间的丰富辩证联系无疑是值得学术界进一步研究和反思的。

参考文献:

- [1] 赵红州. 关于科学家社会年龄问题的研究[J]. 自然辩证法通讯, 1979(4): 29-44. Zhao H Z. The research on the social age of scientists[J]. Journal of Dialectics of Nature, 1979(4): 29-44.
- [2] 梁立明, 赵红州. 科学发现年龄定律是一种威布尔分布[J]. 自然辩证法通讯, 1991, 13(1): 28-36. Liang L M, Zhao H Z. The age law of scientific discoveries is a Weibull Distribution[J]. Journal of Dialectics of Nature, 1991, 13(1): 28-36.
- [3] 赵红州, 梁立明, 王元. 重大科技成果威布尔分布的普遍性[J]. 科学学与科学技术管理, 1992, 13(3): 5-12. Zhao H Z, Liang L M, Wang Y. The Universality of Weibull distribution in major scientific and technological achievements[J]. Science of Science and Management of S. & T, 1992, 13(3): 5-12.
- [4] Liang L M, Zhao H Z, Wang Y et al. Distribution of major scientific and technological achievements in terms of age group—Weibull distribution[J]. Scientometrics, 1996, 36(1): 3-18.
- [5] 夏水华. 对 80 年间诺贝尔物理奖获得者的统计分析及其思考[J]. 科技进步与对策, 2001(3): 128-130. Xia S H. The statistical analysis and reflections on Nobel Prize Winners in Physics over 80 Years[J]. Science & Technology Progress and Policy, 2001(3): 128-130.
- [6] 刘俊婉. 从诺贝尔奖现象看科学创造的特征[J]. 科学学研究, 2009, 27(9): 1289-1297. Liu J W. Some characteristics of scientific creation based on Nobel Prize analysis[J]. Studies in Science of Science, 2009, 27(9): 1289-1297.
- [7] 吴晓东. 诺贝尔生理学或医学奖得主的创造峰值年龄研究[J]. 中国科学院院刊, 2009, 24(6): 655-658. Wu X D. Research on the age of creation peak value of laureates of Nobel Physiology or Medicine Prize[J]. Bulletin of Chinese Academy of Sciences, 2009, 24(6): 655-658.
- [8] 门伟莉, 张志强. 科研创造峰值年龄变化规律研究综述[J]. 科学学研究, 2013, 31(11): 1623-1629. Men W L, Zhang Z Q. Research review of age of peak creative performance[J]. Studies in Science of Science, 2013, 31(11): 1623-1629.
- [9] 樊向伟, 肖仙桃. 科研产出高峰期研究的现状与问题[J]. 图书情报工作, 2015, 59(5): 100-105. Fan X W, Xiao X T. The research status and problems of peak age period of scientific output[J]. Library And Information Service, 2015, 59(5): 100-105.
- [10] Mumford M D. Age and outstanding occupational achievement: Lehman revisited[J]. Journal of Vocational Behavior, 1984, 25(02): 225-244.
- [11] Stephan P G, Levin S E. Age and the Nobel Prize revisited[J]. Scientometrics, 1993, 28(3): 387-399.
- [12] Fortunato S. Growing time lag threatens Nobels[J]. Nature, 2014, 508(7495): 186.
- [13] 王素丽. 基于学科引文的文献生命周期模型探析[J]. 情报理论与实践, 2012, 35(3): 37-41. Wang S L. The analysis of document life cycle model based on disciplinary citations[J]. Information Studies: Theory & Application, 2012, 35(3): 37-41.
- [14] Jones B F. Age and great invention[J]. Review of Economics and Statistics, 2010, 92(1): 1-14.
- [15] Jones B F, Weinberg B A. Age dynamics in scientific creativity[J]. Proceedings of the National Academy of Sciences, 2011, 108(47): 18910-18914.
- [16] Simonton D K. Age and outstanding achievement: What do we know after a century of research? [J]. Psychological Bulletin, 1988, 104(2): 251-267.
- [17] Weinberg B A, Galenson D W. Creative careers: The life cycles of Nobel laureates in economics[R]. NBER Working Paper, 2003.
- [18] 梁立明, 王元, 丁凡, 等. 科学与文化成果年龄分布研究的新发展[J]. 自然辩证法通讯, 1999, 21(4): 28-33. Liang L M, Wang Y, Ding F, et al. New development of age distribution study about science and culture achievements[J]. Journal of Dialectics of Nature, 1999, 21(4): 28-33.
- [19] 蒯亚琼. 人才项目与当代中国学术精英的塑造[J]. 高等教育研究, 2018, 39(11): 1-12. Lin Y Q. Talent projects and the making of academic elites in China[J]. Journal of Higher Education, 2018, 39(11): 1-12.
- [20] Sinatra R, et al. Quantifying the evolution of individual scientific impact[J]. Science, 2016, 354(6312): 596-604.

Further exploration of the peak age of outstanding scientists in scientific research

SHEN Nan¹, WANG Chen²

(1. School of Foreign Languages, Anhui Agricultural University, Hefei 230036, China;

2. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: With the advancement of Internet and information technology, the data generated across various disciplinary domains have witnessed exponential growth. The swift emergence, efficient organization and scientific processing of vast amounts of data have not only become potent instruments but also a novel foundation for academic research. This development has thus given rise to the so-called “data-intensive research paradigm”. The core characteristic of data-intensive research paradigm is its data-driven nature, which means computers are employed to uncover correlations within extensive datasets without the necessity of pre-established hypotheses, enabling the generation of more scientifically rigorous predictive outcomes. This paradigm has opened up new avenues for scientific research, facilitating the re-evaluation of numerous classical propositions on a more expansive empirical scale.

This research focuses on the age aspect within the scientific research life-cycle in the field of scientific sociology, delving deeply into novel conclusions regarding the traditional topic of “the statistical distribution of scientists” peak ages in scientific research. Scholars such as Zhao Hongzhou proposed that the age law of scientific discovery adheres to a Weibull distribution, and Liu Junwan, conversely, based on the ages of significant scientific discoveries of 435 Nobel laureates in natural sciences, put forward a corresponding log-normal distribution. However, in contrast to numerous previous studies, the sampling criteria of the self-constructed dataset in this paper have undergone substantial modifications.

Firstly, the scope of disciplines has been broadened to comprehensively incorporate a wide range of achievements in mathematics, mechanics, geology and computer science. Secondly, considerations have been given to the diversification of scientific elite types. Instead of being confined to scientists solely recognized by top international scientific awards, the scope has been extended to encompass all outstanding scholars who have left an indelible mark on the annals of human science and technology over the past few centuries. Upon conducting software analysis of current dataset, the authors have determined that the optimal fitting function is the “normal inverse Gaussian distribution”.

Further analysis reveals that the optimal fitting distribution of the present dataset exhibits distinct non-static characteristics. Variations in data sampling criteria can lead to disparities in the connotations and extensions of datasets. Regarding the classical topic of “the distribution of scientists’ peak ages in scientific research”, it appears that there is no fixed or static singular conclusion. Given the lack of a unified consensus among scholars regarding the definition of “outstanding scientists”, different datasets established accordingly may well correspond to diverse optimal fitting distributions. Consequently, despite the efforts to gather relevant information through web data mining and manual identification methods, the current dataset inevitably suffers from limitations in terms of sample size and type. In future research, if scholars can design datasets with larger sample sizes and greater sample diversity, it is highly likely that more profound and dynamic conclusions can be unearthed.

In addition to the above, this study also undertakes descriptive analysis and contemplation of local indicators such as the mean age, median age, minimum age, and maximum age of datasets from different epochs. Through observation of recent elections of academicians of Chinese Academy of Sciences, it has been noted that middle-aged and young academicians around the age of 50 are predominantly from the group of “National Science Fund for Distinguished Young Scholars” recipients. This finding further corroborates the distinct connection and isomorphic progression characteristics between “National Science Fund for Distinguished Young Scholars” recipients and “academicians”.

Based on the above, this study finally posits that the concept of “optimal age for scientific discovery” has exerted a profound and nuanced influence on China’s science and technology talent policies. This is evident in the explicit age limits stipulated for applications in various scientific research talent programs across different levels in China, which are set as stringent mandatory criteria. This has led to an increasingly prominent trend of youthfulness and precocity in the selection of scientific elites in China. However, considering the rich individuality and heterogeneous nature exhibited by each scientist during the innovation process, an overemphasis on age restrictions may potentially impede the full realization of individual scientists’ capabilities and runs counter to the intrinsic laws governing scientific and technological innovation. In light of that, the academic community should accord greater importance to the dialectical relationship between statistical regularities and microscopic case studies. While acknowledging the significance of general statistical patterns regarding peak ages in scientific research, it is equally crucial to delve deeply into the rich typical significance embedded within specific scientist cases.

Key words: outstanding scientists; peak age in scientific research; statistical patterns; case study