

传统统计之道，窥数据之妙

——我们的老师方开泰

彭小令 李刚

引言

从20世纪30年代英国学者费希尔（Ronald Fisher）提出了统计推断理论，统计学开始进入蓬勃发展的黄金时代，以数据为核心、研究不确定性的现代统计理论和方法逐渐在各行各业得到了广泛的认可和应用。近几十年，越来越多的华人统计学家开始在国际统计学界崭露头角，统计学在中国也上升为一级学科，得到了前所未有的关注和重视。中国统计学的发展和壮大离不开老一辈开拓者们对统计的深深热爱和一片赤诚之心。这其中举足轻重的一位，就是我们的老师——国际知名统计学家方开泰教授，他就好像天边的明星，为我们这些曾经的统计学子照亮了前进的道路。

方开泰，1940年出生于江苏省泰州市，中学时就读于有名的扬州中学，1957年考取北京大学数学力学系，1963年成为中科院数学所研究生。1967年，方开泰被任命为数学研究所助理研究员，1979年被调到中国科学院应用数学研究所并于1980年10月晋升为副研究员。1980年至1982年，先后在耶鲁大



方开泰在北师港浸大图书馆，2019年

学和斯坦福大学做访问学者。1984年，方开泰被任命为中国科学院应用数学研究所副所长，次年被国务院学位委员会批准成为博士生导师，1986年7月晋升为研究员。1990–2005年间，方开泰一直在香港浸会大学数学系任职，1993年晋升为讲座教授，曾任数学系主任并成立香港浸会大学统计研究与咨询中心，直至荣休。2006年，方开泰接受香港浸会大学吴清辉校长的邀请，来到珠海为新成立的北京师范大学–香港浸会大学联合国际学院（UIC，以下简称北师港浸大）筹建统计学专业。方开泰在北师港浸大每年教授四门本科课程，78岁时才离开一线教学岗位，但至今仍坚持科研工作，笃行不怠。

方开泰教授（以下称方老师）的青年时代正处于新中国曲折成长的历史阶段，他和他的同伴们边学边用，积极投身到统计方法的推广和应用中，取得了丰硕的成果。方老师不但是多元分析的专家，他和王元院士共同提出的均匀设计理论与方法开启了试验设计新的分支，是中国人自己创立的学术领域，方老师也因此获得了“均匀设计之父”的美誉。早在1992年，方老师就被选为美国数理统计学会（IMS）终身会士（Fellow），2001年被美国统计协会（ASA）选为终生会士，并在1995–1999年代表中国任国际统计学会（ISI）理事。2020年恰逢方老师80岁寿辰，《中国科学·数学》特设“庆贺方开泰教授80华诞专辑”，由范剑青和李润泽任特邀编辑，同时，著名的出版社施普林格也出版了书籍 *Contemporary Design of Experiments, Multivariate Analysis and Data Mining—In Celebration of Prof. Kai-Tai Fang's 80th Birthday* 共同祝贺方老师80岁华诞，由范剑青和潘建新主编。该专辑和著作介绍了方老师在试验设计、多元统计分析、广义多元统计分析和数据挖掘等方面的杰出成就。



本文的两位作者，曾在不同时期跟随方老师学习和工作，师恩难忘。谨以此文回顾方老师对中国统计学的发展和推广，对统计学学科建设和统计人才培养等方面的贡献和成果。

(一) 在实际应用中发展统计学

从中科院学习运筹学的研究生到一代统计大师，方老师不只是在书本上，更是在一个个的实际项目中逐渐领悟统计思想，在应用统计方法的过程中又提炼出新的统计研究课题，从而不断扩展统计学，并开创了均匀试验设计这一统计学新领域。

1.1 普及正交试验设计，开创均匀试验设计

由于新中国工农业发展的迫切需要，1971-1975年，方老师积极投入到正交试验设计的普及活动中。他发明了正交试验设计的“直观分析法”，通过图表的方式使工程师更容易理解正交试验设计的思想和数据分析¹。1976年，中科院数学研究所指定方老师负责中央电视台一个报道和宣传推广正交设计的节目，为了让统计方法能被更多的观众接受理解，他花了大量的时间准备道具和讲解稿。最终，中央电视台在晚间“新闻联播”后的黄金时间足足用了17分钟来介绍正交试验设计。因为准备充分，节目达到了很好的推广效果。



方开泰教授和王元院士，中央电视台，1994年

1978年，原七机部三院三部（现航天总公司三院三部）有三个项目在研制导弹指挥仪的数学模型时遇到难题，由于试验次数有限，不可能按照正交试验设计的要求来实施。为了解决这个难题，方老师和中科院的数学家王元决定运用数论知识建立一个全新的电脑仿真试验设计的理论和方法。那段时间，每天一大早王元就到方老师宿舍楼下叫上他一起去办公室讨论，沉浸在研究中的两位数学家忙碌又充实。这个新的理论和方法就是均匀试验设计法^{2,3}，使用均匀设计不仅顺利达成了项目预期，更是开创了试验设计新的分支。因为均匀试验设计既可以用于电脑仿真试验，也可以用于传统的实体试验，对模型有稳

¹ 方开泰, 刘卓雄 (以集体名义). 正交试验设计. 有色金属, 1974, 8: 39-56.

² 方开泰. 均匀设计——数论方法在试验设计的应用. 应用数学学报, 1980, 3: 363-372.

³ Wang, Y, Fang, K T. A note on uniform distribution and experimental design. Kexue Tongbao (Chinese Science Bulletin), 1981, 26: 485-489.

健性，试验次数灵活，很快在航天、化工、制药、汽车制造、科学研究等领域得到广泛应用。之后，方老师和他的研究团队发展完善了均匀试验设计的理论，该项目获得 2008 年国家自然科学二等奖（是年一等奖空缺）。



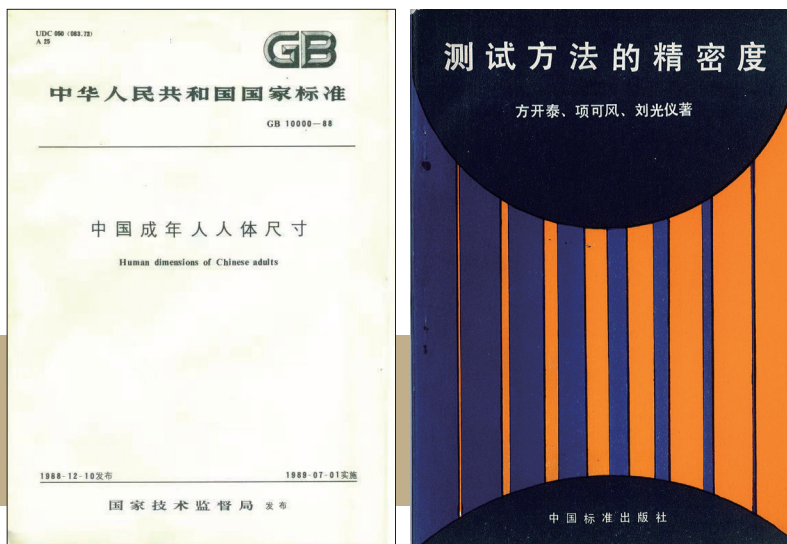
方开泰和刘永才（我国第一代陆基巡航导弹总设计师，2008 年国家科技进步特等奖获得者）



方开泰和王元在 2008 年度国家科学技术奖励大会

1.2 统计方法与国家标准

在新中国发展的进程中，许多行业需要制定国家标准，其中统计学家功劳不小。方老师在二十世纪 70 年代先后参与了审核合金结构钢国家标准、制定中国成年人服装标准和测试方法精密度标准三项国家级项目，根据实际情况采用了不同的统计方法，最终这些看上去与统计毫无关系，甚至让专家也不知所措的行业标准问题，却依靠统计学得到了完美的解释和解决。



1973年，方老师和中科院数学所同事吴传义接受了冶金部的一个项目，审核合金结构钢国家标准。所谓合金结构钢，是在炼钢时控制碳、铬、锰、镍、钼、硅等元素使其含量必须落入国标规定的范围，而炼出的钢材需要检验其五个机械性能、强度、弹性等是否合格。全国有十多个工厂生产同一种合金结构钢，如北京钢厂、抚顺钢厂、大冶钢厂、齐齐哈尔钢厂等。他们发现，即使化学元素碳、铬、锰、镍、钼、硅等完全符合国标，也不能保证钢的机械性能全部合格。如齐齐哈尔钢厂，合格率才达38%。众所周知，合金结构钢十分昂贵，如果不合格，生产厂将承受巨大的经济损失，于是，不少钢厂认为国标有问题。而该国标是解放初从苏联引进的，谁也不知其原理。冶金部请北京钢铁研究院来鉴定国标的合理性，他们觉得该项目难度很大，便请数学所一起攻关。为了搞清楚问题，冶金部从全国有关钢厂调来数据，工作团队便开始了长达半年的数据清理工作，如根据实际情况剔除一些不属于研究总体的经过特殊处理的数据。最终，方老师和吴传义建立回归分析模型，通过钢的化学元素来预测钢的五个机械性能并通过计算五重积分获得钢材合格率。数据分析的最终结论为(i) 国标是科学的、合理的；(ii) 钢中不同元素组合，相应的机械性能合格率不同，因此，在炼钢过程中，要尽量向合格率高的元素组合靠拢；(iii) 北钢和抚钢的“土法”经验是对的。这个项目从大量的实际数据出发，经历了数据清洗、建模最终得出了让人信服的结论，可算作早期的“数据挖掘 (Data mining)”。从项目中方老师还提出了一个新的统计理论问题⁴。这个理论问题直到2020年⁵才圆满解决。

⁴ 方开泰, 吴传义. 一个概率极值问题. 应用数学学报, 1979, 2:132-148.

⁵ Fang, K.T., Luo, Z. and Tong, Y.L. (2020), Convexity of sets under normal distribution in the structural alloy steel standard, in Chapter 3 of Festschrift in honor of Professor Dietrich von Rosen on his 65th birthday, 41-49, Springer.