

方积乾教授在 2024 年中国生物统计学术会议上的讲话

方积乾

(2024 年 7 月 20 日)

[编者按]2024 年 7 月 19—21 日,中国生物统计学术年会在广州市召开,来自海内外的 1000 余名生物统计专业人士参加了会议。中山大学国家教学名师方积乾教授作为生物统计学领域前辈,应大会特邀在开幕式上对青年学者寄语。方教授提出“三学习”——数学、现代计算技术、服务对象,“三传统”——尊师爱徒、合作共赢、同行厚道。85 岁高龄的方教授激情洋溢的演讲和谆谆教诲赢得了全场参会人员的热列掌声。应广大参会者和读者要求,我们将方教授的讲话内容整理成稿,以飨读者。

DOI 10.11783/j.issn.1002-3674.2024.05.001

各位,我今天非常高兴,1000 多位参会人员齐聚一堂,个个报告都精彩,个个都是人才。我一边听一边向周围的人员打听,这个报告人很不错,在哪里工作,我就想把这些人才都挖掘到我自己的单位来。

大会组委会要我给大家讲几句话,我有一些想法与大家共勉。

第一个想法来自于我们学科的特点。我们这个学科英文的定义是:Biostatistics is a discipline that is concerned with the development and application of statistical theory and methods for the study of phenomena arising in the life sciences。这是我的导师蒋庆琅先生给的定义,我把它翻译成中文:“生物统计学是一门发展和应用统计学理论和方法的学科,研究生命科学中发生的现象”。注意是发展,不单是应用。刚才林希虹教授的报告就是既发展又应用,是一个典范。

再看关于统计学的定义:Statistics is the science and art of dealing with variation in data through collection, classification, and analysis in such a way as to obtain reliable results。这个来自 Webster 词典的定义将统计学看成是科学与艺术。注意没有?既是科学又是艺术,其他学科很少是这样的。

为了学习统计学新成果,发展统计学理论与方法,我的导师蒋庆琅先生号召“mathematics, study mathematics, take more courses in mathematics”。

在新冠流行期间,大家都有做核酸检测的经历。我们将 10 个人的样品合并成 1 管,然后送去测定,叫做“混管测定”。这是蒋庆琅先生当年用了一点点数学知识发明的,服务于当时流行病学研究蚊子病毒感染率的问题。

当时有人收集了 500 只母蚊子,要测定这些蚊子被某种病毒感染的百分比,一个一个蚊子去测定的话,要测定 500 次。感染率很低,检测费也很贵,500 个都测,就很浪费。

蒋庆琅先生提出把 500 只蚊子分成 $n=20$ 组,每

组 $m=25$ 只。然后,20 个组,一组测一次,一共检测 20 次。假如有 x 组感染了,用 x 除以 n ,就得到一组蚊子被感染的概率。基于此,就可以用数学公式来计算那个感染率了。原理是什么呢?一点点数学和概率论的知识。一只蚊子被感染的概率如果是 P ,未被感染的概率就是 $1-P$,那么一组内 m 个蚊子没有被感染的概率就是 $1-P$ 的 m 次方,一组蚊子被感染的概率就是 1 减去它,于是就得到了一个等式: $x/n = 1 - (1-P)^m$ 。从这个等式就可以解出 P 来了。当时这个方法解决了 20 世纪 70 年代的蚊子问题,后来文献报道,20 世纪 90 年代又将这个方法用到 HIV 感染率的测定中,而前几年用到了我们新冠病毒感染率的测定上了。可见得,这一点点数学就发挥了作用,所以,我们要学习 mathematics。

第二,要学习现代计算技术。过去我们计算是用计算尺,用计算器,用手摇计算器,后来有了 computer,现在我们要学习现代的计算技术。

我们大家知道,20 世纪 70 年代的时候,经验分布加上计算技术就产生了一个统计学里边的新方法 bootstrap。判别分析是我们统计学里的方法,加上计算技术会是什么?就是有监督的机器学习;聚类分析加上计算技术产生什么?就是无监督的机器学习。现在很热门的机器学习无非是统计学和现代计算技术结合在一起。决策树是我们统计学的方法,决策树是 20 世纪 80 年代伯克利教授提出来的,加上计算技术是什么?就是决策森林。相关系数是很简单的统计学方法,相关系数加上计算技术得到什么?得到了一门学科——网络药理学,大家听说过没有?就是相关系数构建了药理的网络。统计学加计算技术是什么?人工智能。

今天到处都讲的那个 AI,无非是把统计学的思想与方法与现代计算技术相结合。必须说明,在座的统计学家,我不要求你们改行,而是要求你们一定要学习现代计算技术;并不是把你们都变成 AI 专家,把我们统计学消灭了。我们要扬长补短,统计学之长是什么?模型驱动,我们有一套统计学理论;统计学之短是什

么? 缺乏大数据计算技术, 所以我们要深入研究大数据的统计学理论、方法和计算机实现。今天林教授的报告就是个典范, 她研究了大数据的统计学理论、方法和计算机实现吗。最后一定要计算机实现, 要提供软件, 统计学加大型计算技术等于数据科学。大家听说过吗? 现在有数据科学了, 山东大学建立了数据科学的专业, 开始招收本科生。

第三是学习服务对象。我们统计检验加临床实践, 那是什么? 序贯检验。大家记得我们统计学里的序贯检验、序贯思想, 就从临床看病学来的。判别分析加临床实践, 产生了判别树。判别树不是一下子就判别一个结果, 而是逐渐的、一步一步来的, 就像我们看病的实践, 先量体温, 发烧不发烧, 分为两批, 发烧的人再去查肺部, 再去查 CT 等等。RCT 大家都熟悉, 随机对照试验加我们几十年的临床实践, 现在发展成什么? 适应性设计。你 RCT 要求患者吃一种药, 一下子一竿子到底, 不符合临床实际, 治疗一段时间以后看效果, 必要时改变我的治疗策略。大家学过工具变量 instrumental variable, 这是在经济统计里常用的方法, 和 GWAS 结合起来产生了什么? 产生了孟德尔随机化, 天生的老天爷给我们的随机化方法。统计学加分子生物学, 产生了什么? 生物信息学。刚才林教授坐在我旁边, 我问她“你那些生物学知识哪里来的?” 她说, “我 10 多年前自学的, 学了以后也觉得不难了。” 总之, 我们统计学家一定要去学你服务对象的理论。

现在说说我本人的做法。我在医学院里边长大的, 22 岁就到医学院教书, 一直到现在 85 岁。第一, 要下医学研究现场, 学对方的专业文献, 我最早是下针刺麻醉的现场。我在针刺麻醉的研究组里边, 一方面帮他们解决统计问题, 另外一方面学习医学知识。

我现在发展到喜好医学文献, 我看的医学文献是那些专业医学工作者都来不及看的, 他们忙, 没时间看, 我给他们讲课的时候都讲他们的 2021 年、2022 年的文献。我非常耐心地去听取他们的故事, 理解了以后, 我就设身处地地给他们献计献策, 来“推销”统计学的应用。学了他们的知识, 开动脑筋再给他们献计献策。我最早这样做是什么时候? 就是在 20 世纪 60 年代初文革前, 那时候超声诊断刚刚进到中国。超声诊断的创始人张武, 那时是我们北医的研究生, 他来问我微积分里边面积测量的事情等等, 我就知道了面积测量原来在超声里那么重要, 所以我学了他的东西以后, 又向他推销面积测量, 建议他测哪几个参数可以近似算得测量面积。

我的研究课题很多都是向医生们学习产生的。

首先就是针刺麻醉的数据分析。大家知道生存分析里边不是有删失 censoring 吗? 我最早就是在针刺麻醉研究中应用了 censoring。针刺麻醉的研究, 用电炉丝去烧大鼠的尾巴来测定其“痛阈”, 烧大鼠尾巴时间一长, 大鼠忍不住痛, 就甩尾巴了。在第几秒钟甩尾

巴? 记下来, 就是它的痛阈。但是, 有的大鼠非常坚强, 烧到 30 秒还不甩尾巴, 再烧下去要烧焦了, 于是停止, 那么大鼠的痛阈是多少? 就是 30+, 这就是 right censoring 右删失。

还有药物有效期预测。你们知道药物有效期怎么测的吗? 因为药物摆在柜子里, 它 20 年还有效, 你怎么能得到有效期? 要采用“加速实验”, 用到常微分方程。肿瘤自然史涉及到“多状态随机过程”, 我的导师是多状态生存分析的创始人, 我的博士论文就是“具有时变协变量的多状态生存分析”。我们比较早在北京就做了电子计算机的中医专家系统。还有离子通道门控动力学, 我本来不知道离子通道的, 有人来问我, 原来离子通道的信号就是脉冲信号, 脉冲信号就有密集的和稀疏的, 那就用到我们随机过程中的点过程。

凡此等等, 在向服务对象学习的过程中, 产生了自己研究的课题。

今天我要讲的第二部分是发扬“三传统”。

今天这么多人来参会, 七个学会在一起开会, 就是我们发扬传统的结果。

第一个传统是“尊师爱徒, 世代相传”。解放初的时候, 我们中国只有 4 位统计学导师, 大家要记住他们的名字, 以年龄为序, 分别是上海医学院的许世瑾; 薛仲三, 最早是复旦大学的, 后来在军事医学科学院; 预防医学科学院的李光荫; 第四军医大学的郭祖超, 他最年轻, 解放后为国家工作了 50 年, 所以大家比较熟悉郭祖超。但是前 3 位我们也不能忘记。这 4 位导师在解放初的时候办班、招进修生, 培养出了一批学科带头人, 这些人都是谁? 都是在座各位的导师, 都是你们所在单位的学科带头人。看我们今天在座的第一排, 王彤, 他的导师是何大卫; 夏结来, 导师郭祖超; 陈峰, 导师陆守曾, 等等。这些人我们不能忘记, 他们都是解放初四大导师培养出来的学科带头人, 而正是这些学科带头人, 培养出我们今天的骨干。

这张照片是 1982 年卫生部举办的寿命表培训班合影, 刚才说的那些骨干大多数在这张照片里面。1982 年, 蒋庆琅教授来讲寿命表, 我有幸当助教。



图 1 卫生部及 WHO 举办“卫生统计师资讲习班”合影(1982 年 8 月)

(下转第 649 页)