



统计百家丛书

陆守曾文集

陆守曾 著



中国统计出版社
China Statistics Press



统计百家丛书

陆守曾文集

陆守曾 著



中国统计出版社

China Statistics Press

图书在版编目(CIP)数据

陆守曾文集 / 陆守曾著. — 北京 : 中国统计出版社, 2016.11

ISBN 978—7—5037—8034—9

I. ①陆… II. ①陆… III. ①医学统计—文集②卫生统计—文集 IV. ①R195.1—53

中国版本图书馆 CIP 数据核字(2016)第 255022 号

陆守曾文集

作 者/陆守曾

责任编辑/杨映霜

封面设计/张 冰

出版发行/中国统计出版社

通信地址/北京市丰台区西三环南路甲 6 号 邮政编码/100073

电 话/邮购(010)63376909 书店(010)68783171

网 址/<http://www.zgtjcs.com/>

印 刷/河北鑫宏源印刷包装有限责任公司

经 销/新华书店

开 本/710×1000mm 1/16

字 数/310 千字

印 张/19

版 别/2016 年 11 月第 1 版

版 次/2016 年 11 月第 1 次印刷

定 价/49.00 元

版权所有。未经许可,本书的任何部分不得以任何方式在世界任何地区以任何文字翻印、仿制或转载。

中国统计版图书,如有印装错误,本社发行部负责调换。

献 给

黎宗琦女士

编辑说明

改革开放以来，我社每年都有一些统计论著问世，这些统计论著的出版，为我国的统计理论建设作出了积极贡献，取得了良好的社会效益。但是，由于这些统计论著一直是每年分散出版，再加上时间的因素，使得这些富有思想性的论著不利于读者的系统收藏和整理。为了改变这种状况，我们特设置“统计百家丛书”，专门收集出版统计论著方面的优秀著作，以使这些统计优秀著作以一种整体形象出现，形成规模，这样既便于读者阅读和使用，又便于系统保存，为后来者留下系统的参考图书。这是我们编辑出版“统计百家丛书”的第一个目的。

任何事业的兴旺发达都与它的理论发展与创新密不可分，从这个意义上说，统计事业的兴旺发达，也有赖于统计理论的发展与创新。发展与创新是新世纪赋予我们的历史责任，作为为统计事业服务的出版者，我们自觉负有促进统计理论繁荣、培育统计理论研究人才的责任。所以，“统计百家丛书”是为繁荣我国的统计理论建设而创立的，它的宗旨就是为我国广大的有志于统计理论研究的人员提供一个言论的舞台。一切热爱统计事业、并为统计事业的发展刻苦思考、不断探索的有志之士，都有可能把自己的一家之言在这个舞台上公之于众。这是我们编辑出版“统计百家丛书”的第二个目的。

我们不谋求“统计百家丛书”中的每一本论著都能高屋建瓴，或者是鸿篇巨制，但是，我们可以保证，在“统计百家丛书”中，每一本都是作者热爱统计事业，并为统计事业而艰苦努力的呕心沥血的拳拳之作。它们或者是对改进统计方法的创新性探讨，或者是对改革统计制度的大胆直言，或者是两者兼而有之。无论如何，它们所有的内容都将见证中国统计事业的进步与发展，并为中国的统计事业留下宝贵的理论财富。这是我们编辑出版“统计百家丛书”的第三个目的。

如果“统计百家丛书”得到了您的喜爱，并使您开卷有益的话，那么，我们的第四个目的就达到了。希望更多的人来关注“统计百家丛书”！

“统计百家丛书”编辑部

我选择医学统计学为终身事业，已经 64 年。当初，毅然的决定有什么机缘？有什么原因？后来，又有什么条件使我始终执着不懈？就以对这方面的回忆作为本《文集》的“前言”吧。

（一）1952 年初的某一天，南京，在原中央大学医学院公共卫生科的一间小会议室里，郭祖超教授为我们 5 名公共卫生专业本科生讲授医学统计学。这个会议室中央放着一张长条桌，我们 5 个人被指定分坐两侧，郭教授站在条桌的一头，他后面横立着一块大约 2 米 × 1 米的黑板。在我们背后，各摆开一排杂样的凳子或椅子，已经端坐着十余位旁听者；尽管在我们身边还有好几把与长条桌配套的椅子空着，却无人来坐，可能是在旁听者和我们学生之间有一条无形的“线”，他们不便越过。郭教授说的第一段话大意是：这门课程专为这 5 位学生开设，共 120 学时，其中包括课堂实习；你们可以在课中提问，但必须在我讲完一个段落时。各位旁听者就不要提问了，不能影响他们的学习。接着，郭教授开讲了。我意想不到的，就从这一刻起，我竟然完全被斯人、斯时、斯地的一切深深地、牢牢地吸引住，被带进一个从未见识过的、妙不可言的境界。这就是我此生第一次聆听名师讲授医学统计学。如今已过去 64 年，我实实在在地未曾离开本专业半步！同时，郭祖超教授当然就成为我最崇敬的恩师，在我毕生的事业中，他始终引领、激励、推动我前行。回顾这段往事，那是我刚刚步入大学四年级下学期，正彷徨于今后的出路乃至一生志向的关键时刻。这个突然降临的幸运一下拨开了我眼前的迷雾，我义无反顾地全心投入医学统计学事业中。

（二）在上世纪 50 年代初，医学统计学是个冷门学科，难见其发展前景，我却对它如此痴迷，有两个原因：一是我本钟爱数学。原以为进了医学院，就事实上了断了同它的亲密，虽不舍，也只能割爱。而医学统计学的出现，不但使数学这个“尤物”又回归到我身边，而且拥有这份强势的助力，更增添了我投入本专业的信心。二是医学统计学的特殊魅力吸引了我。嗣后随着学习的步步深入，我逐渐“触摸”到它的真谛：既抽象又具体；既严格又灵活；既深邃又显而易见；既是客观存在，又包含主观判断；等等。在医学领域近百个学科当中，医学统计学没有生物学的形态和功能，也不能用来治病，

尤如一门“另类”，却正中我的喜爱。

（三）在医学领域，如果有人想用一用医学统计学，来帮助自己完成所从事专业的某项科研或其他工作，那么，看点书，或听点相关的课，一般就可以解决了。但是，如果要选择它作为长期从事的专业，那是有条件的。几个？两个半。第一，喜欢它。只要细心体验，统计数据是有“灵性”的。它们的出现和分布既有变异性，又有特异性；抽样误差作为随机现象，看似毫无章法，其实始终遵循着某种规律。经常与这样的对象为伴，观察它，欣赏它，摆弄它，分析它，从中提炼出信息，寻找其规律，再作出判断，当然就充满乐趣。第二，有悟性。医学统计学面向的是医学范畴的问题，又同时遵循统计学原理，它把两者结合而形成自己的“个性”。从业者必须靠自己的悟性和努力来理解它，各人的结果可能是不同的，别人难以帮助。有人虽然十分喜爱本专业，也确实下过苦功，但最终遗憾地未得其门而入。第三，在应用领域拥有一个或几个项目，认准其科学价值，投入进去，锲而不舍，数年或数十年，取得成果。我曾有机会以医学统计学的优势投入地方病研究、中西医结合研究、解放军卫生工作研究，每个项目均达数年或数十年，并取得了一些成果；此外，还涉猎过体育统计、教育统计等方面的项目。然而，由于项目众多，看似琳琅满目，实则杂乱无序，又互不关联，亦并非我之初衷。须知在当年，囿于历史原因，个人自主选择项目的空间极为有限，大多是被授予什么任务，就完成什么项目。因此，从业者如果由于主观或客观上的原因，在这方面未能如愿以偿，也仍然可以在本学科达到较高的造诣。所以，这只能算得上“半个条件”，亦仅是我的建议而已。

我已进入耄耋之年，在风云变幻的几十年里，有幸始终沐浴在我酷爱专业的光耀之中，不断汲取到知识的精髓。我愿把所历、所为、所思、所悟奉献予同行学者，尤其是青年学者。

本“文集”收录了本人1955～2012年间发表的论文、综述、学术研讨、科普文章、短篇、杂文共48篇。请读者批评、指正。由于时间跨越半个多世纪，早期文章中一些专业名词与现在不一样，例如：“显著性测验”对应于“假设检验”，“有显著性”对应于“有统计学意义”，“全体”即“总体”，“团体设计”即“平行组设计”，“变异数分析”即“方差分析”，“典型抽样”即“分层抽样”等。为了真实反映当时学术界的传承和通用状况，本次编辑出版时保持了原样，未做任何修改。

我衷心感谢陈峰教授和他的优秀团队为本“文集”的出版所做的一切！感谢肖静、荀鹏程二位老师的支持和帮助！

陆守曾谨识

2016年4月30日于南通

1. 实验设计中关于统计方面的几个问题 / 1
2. 长春市 1955 年和 1958 年男女人口简略寿命表 / 15
3. 中国海员截缩简略寿命表 (1979-1983 年) / 26
4. 浅谈人口平均寿命 / 32
5. 长春市 1955 ~ 1958 年人口出生率的季节变动 / 33
6. 身体发育调查中关于年龄分组的商榷 / 35
7. 用常态相关面评价个体身体发育的探讨 / 40
8. 计量资料原记录准确性的统计检查法 / 49
9. 变异数的显著性测验及其相应的平均数显著性测验 / 53
10. 抚松镇克山病发病率的统计分析 / 60
11. 有关克山病流行病学方面几个问题的商讨 / 68
12. 克山病的主要流行规律和病因线索 / 82
13. 对克山病主要流行规律的看法 / 89
14. 吉林省克山病统计资料初步分析 / 99
15. 吉林省克山病 25 年资料统计分析 / 113
16. 《中华人们共和国地方病与环境图集》编纂中的
调查设计与统计分析 / 123
17. 有关医用统计学方面几个问题的商榷 / 130
18. 医学应用数理统计的几个问题 / 133
19. 几种非参数统计法 / 139
20. 显著性检验中的 I、II 型错误 / 155
21. 如何选用统计方法 / 159
22. 四格表的确切概率工具表及其计算方法 / 168

23. 四格表显著性界值表的改进与补充 / 178
24. 六种检验四格表资料显著性方法的评价 / 186
25. 对四格表资料六种检验方法的再评价 / 191
26. 医学杂志中正确应用统计方法的几个问题 / 193
27. 中医药科研设计中的交互作用和均衡性 / 198
28. 中医药科研设计中应用统计方法的几个常见问题 / 201
29. 统计资料的综合分析 / 207
30. 两条回归线高度的比较 / 211
31. 怎样正确应用假设检验 / 219
32. 2×2 析因设计及其分析一例 / 222
33. 对估计样本含量的认识 / 226
34. distribution-free 与 nonparametric 有何区别? 有些书刊中以
distribution-free 为题, 却阐述非参数检验的内容, 应如何理解 / 228
35. 对《统计数值表》一个界值的改正 / 230
36. 医学应用变量变换法质疑与建议 / 231
37. 现有地方病统计资料再开发中的几个常见问题 / 238
38. 对“再开发”研究的再认识 / 240
39. 现有资料再开发研究 10 年 / 242
40. 对卫生统计学的性质等两个问题的看法 / 246
41. 均数比较的常见误用 / 251
42. 单侧检验辨释 / 253
43. 回顾与前瞻 / 260
44. 从宏观视角认识医学统计学 / 270
45. 对医学统计学应用现状的四点看法 / 276
46. 我对培养专业人才的认识 / 279
47. 郭祖超传 / 281
48. 师恩永恒 传承无限 / 290

1 实验设计中关于统计方面的几个问题

由于研究工作的开展,对研究中有关统计资料的处理方法也就逐渐感到需要。无论从实验的设计、资料的收集、整理和分析以致一张表的设计和一幅图的绘制等若能适当地采用一些必要的统计方法,对研究工作的计划、进行和所得结果的解释都是有益的,而统计方法在研究工作中的应用首先就在于实验设计方面。当然,有的研究工作需要统计处理的,有些是不需要的,这里所讨论的实验设计以前者为限。至于采用各种实验设计所得结果的统计分析方法则不在本文讨论范围之内。

在需要统计处理的研究中运用了合理的实验设计将会有助于获得预期的效果,得出正确的结论;而由于设计方法的不周密或不合理,小则丧失一部分有用的知识,大则不能得到正确的结论以致损毁了全部实验而完全得不到结果。例如,有一个实验要比较一种间接的测定维生素丙的方法是否可代替标准测定法(二种)而得到同样正确的结果。我们从同一样品中取出同样的样本若干个,由甲、乙、丙三位实验者分别用二种标准方法与一种简捷方法各重复测定若干次以观察其区别与误差,安排如下:

实验者	方 法	次 数
甲	第一种标准测定法	1 2 3.....N
乙	第二种标准测定法	1 2 3.....N
丙	简 捷 测 定 法	1 2 3.....N

以上排列表示实验者甲用第一种标准测定法重复操作 1,2,3,.....至 N 次,实验者乙或丙亦各用第二种标准测定法或简捷测定法重复操作 N 次。如果简捷测定法所得到的维生素丙含量的平均数较之二种标准方法的结果相差很大(即在统计上相差显著),那么是否就可以认为简捷测定法不可靠呢?这个结论还是值得商榷的。因为实验者甲、乙、丙三人的技术条件对三种测定方法各自起着不同的影响,因为没有二个人所得到的结果会完全相同,这种人为的因素渗进了实验的结果但未被区别开来,假定这三种方法的结果本来并无

真正的差别,仅由于实验者丙的技术水平较差以致结果不准确,这样就误把实验者之间技术水平的相差当作了简捷测定法和标准测定法之间的相差,其结论自然就不正确了。比较合理的实验设计应该这样安排:

方 法	每次实验者						
	甲	乙	丙	甲	乙	丙	
第一种标准测定法	1	2	3	4	5	6	N
第二种标准测定法	1	2	3	4	5	6	N
简捷测定法	1	2	3	4	5	6	N

上面的排列表示实验者甲、乙、丙同时用三种方法操作并重复若干次,这样,实验者技术条件不一致的因素将同样加诸各种测定法并与各种方法间的相差分别开来,三种方法所得的结果也就真正能够代表他们的优劣了,如果简捷法一组所得的平均数和另二组之间的差别仍然很大(相差显著)则可以认为简捷法尚不甚可靠;而当三组结果一致(相差不显著)时即可认为简捷测定法与标准测定法所得结果并无甚出入而可以以之代替了,由此可见实验设计在研究工作中的重要。后一种设计方法是根据“随机区组”来的,下面第三节还要讨论。

现就需要统计处理的医学研究工作中有关小样本方面的实验设计提出以下四个问题以供参考,并请指正。

一、全体、样本与抽样方法

全体是研究的全部对象,而实际进行调查研究或实验室中观察的动物则往往只是这个全体中的一部分,也就是样本。真正的全体有时是完全假想的,例如研究一种新疗法的疗效,开始时只试用于极少数的病人,但其目的显然是要使这种新疗法适用于所有同样性质的病人,因此,这个假想中的全体便包括了所有接受这种新疗法治的全部病人(假使研究成功的话),像这样的全体是当时无法确定其数目的;另外也有可能确定的全体,例如调查某地钉螺体内日本血吸虫的自然感染百分比,这个真正的感染百分比是当时存在的,通过当地某一时间内对全部钉螺的调查是可以获得的,但在一定的人力、物力和时间的条件下,往往不易做到或并无必要,而通过对一定数量钉螺的调查同样可以得到在实用上有价值的数字(自然感染百分比)。至于象实验室中的小白鼠、蚊虫密度、人体的血像检查、饮水水质的检查等等则只是用样本来对全体作出若干推论了。既然对样本的研究是为了说明全体,那么在取样以前必须对希

望了解的全体定出明确的界限,当样本被总结以后,研究者必须审慎地对实际被抽样的全体做出结论,而只有当所有样本能够代表全体,成为整体的一个“具体而微”的“缩影”时,这种结论才有价值。

为了使样本能够代表全体,取样时应遵循以下两个条件。

(1)对各个单位的选择应严格客观地进行,要排除任何成见,亦即按照随机的原则抽取。有些有经验的研究者,他对全体具有亲切的知识,他所选择的样本可能对全体作出良好的估计,但这种“抽样”的危险性远比机遇因素能自由作用者为大,因为这种主观的人为的选择往往易于产生偏性特别在变异度方面常较实际情形为小^[1]。随机的意义在于避免偏性,而有偏性的样本是不适于代表全体的。如像在打靶实习时瞄准的方向都向着红心,但由于初学者的技术不熟练,子弹往往打在红心的四周,当射击许多次以后,可以见到弹痕是向各个方向散布开来而仍以红心作为中心趋势的,这就是一种无偏性的散布,虽然不是每次都中红心,但多次射击的弹痕仍可对红心的位置作出良好的估计。可是当枪筒本身不正时,则即使每次射击都瞄准的红心,而弹痕的散布必趋向于另一方向,这就是由枪筒不正所造成的“偏性”,这时弹痕所围绕的便不是红心了,若用这些弹痕所围绕的中心来估计红心的位置便不会得到正确的结果。随机取得的样本(无偏性的样本)可以运用机率的原理以显著性测验或可信限等统计方法来阐明其结论的误差范围,而不是随机取得的样本则不能运用这些方法,其对全体估计的可靠程度自亦无法推测了。

(2)被抽样的单位应足够的多,这是要求样本的大小应该合宜,过小的样本是不能用来代表全体的。关于这一点在第二节中还要详细讨论。

常用的抽样方法有以下三种:

(1)随机抽样——被研究的单位是由整个全体中随机选出的,在这全体中的每一个单位都有同等被选取的机会。例如自然界蚊虫密度的估计和血液常规检查中的皮肤取血都应该是在随机状态下从全体中抽取的一小部分,这就是随机样本。但当捕蚊者技术水平相差悬殊时各种密度便不能作客观地比较,而取血时若已刺破皮肤再用力挤压所得的血液其检查结果也就不能完全代表全身血液的实际状况了,因为前者使每一蚊虫落到某一样本中的机会并不均等,而后者因血浆被挤出较多而使血液稀薄,这是在随机取样时所必须注意的。

(2)机械抽样——按照一定的机械的顺序抽取被调查的单位,而各单位的排列或散布必须与这种机械的顺序完全无关。常用的机械抽样有目次法和棋盘法。目次法例如某地调查日本血吸虫病的流行情况时将抽查村按户籍每十户抽查一户^[2],因为公安部门的户籍顺序是与日本血吸虫病的流行无关的,这

便使得到十分之一户口数的机械抽样的样本。棋盘法如血球计数,因为血球在计数盘上的散布是随机的,而盘上的方格则机械地割分,二者无关,这样也就得到了随机的样本。另如流行病调查中利用城市中的街道将地区划成方格,对于按机械方法规定的一定数目的方格内的区域则进行调查亦是棋盘法的例子。

(3)典型抽样(亦称分层抽样)——预先将被研究的全体分成若干典型组,然后再在各组内进行随机或机械抽样。例如调查某地蚊虫密度时,各蚊虫在城区和郊区的分布既不同,而野外树林和牛棚猪圈中各个时间的密度变化亦相当大,幼虫在稻田中的散布甚至中央和边缘亦不一致,这样,调查时就必须先对各种类型的地区分成组,再进行抽样,若需要一些地区作对照时亦应适当地选择。曾有人在实验“六六六”的杀虫效能时以城区作郊区的对照,这样的实验设计即使在喷洒“六六六”后郊区的蚊虫密度显示下降亦是不足以说明其功效的,因为失去了合理的比较,这是企图把不同质的两个样本的结果混淆在同一个全体内的设计方法,当然是极不合理的。

二、样本大小之确定及其条件

随机的抽样方法和样本的适当大小是保障样本代表性的两个重要条件。抽样方法不合宜则不易得到无偏性的样本,而过小的样本也同样不能对全体作出有价值的推论。例如有一个研究几种中草药镇咳作用的实验^[3],该实验用猫 39 只分八组(包括对照组)进行,最多的一组用 11 只,最少的两组各仅一只。从统计上看,用一头动物的组是不能作出结论的(若是典型的一例,则为另一问题,它不属于统计处理的范围,但若实验时亦仅用一头动物亦无从决定其是否典型)。因为动物个体间的变动很大,实验中所产生的误差是不能从一头动物计算出来的,因而也就无法测定其可靠程度,故原文对仅用一只猫的实验组所作的结论是尚待商榷的。其实这个实验在设计时若每组(包括实验组)都用猫五只则总数为 40 只,只比原用之 39 只增加一只,但所得的结论要可靠的多了。可见合理地运用实验设计方法并不是一定要很高的条件,往往可以利用原来的仅有条件作若干适当安排就可能得到较好的效果。至于样本究竟要大到什么程度才合适,是否越大越好,这就必须对全体的变异(如标准差)先略有知识,否则关于样本大小的确切数字是无法凭空决定的,应参考过去文献所载,或由初步实验的结果提供若干资料方可据以作出合理的估计,常用的估计样本大小的方法如下:

(1)测量资料中样本大小的估计——有许多量数如身高、体重、胸围、身体

表面积、体温、药物剂量、营养素含量、摄入热量等都是测量资料,另如呼吸次数、脉搏次数、血球计数、昆虫密度等都是计数的,一般来说其原始资料中没有小数,但这类资料的性质和特点都与测量资料相似,这些资料在一定同质的条件下都可以求平均数和标准差,处理方法亦相同,其样本大小的估计可用下式^[4]:

$$N = \left(\frac{ts}{x} \right)^2$$

$[t]$ 为常态曲线下与一定机率相当之界限,若需使结果精确至 95% 则用 $t = 1.96$,精确至 99% 则用 $t = 2.57$,可参考常态曲线机率表^[5], $[s]$ 为标准差, $[x]$ 为样本平均数与全体平均数间之差数, $[N]$ 即样本应有之大小。如某营养实验以鼠 9 头均先后饲以黄豆鸡蛋粉与全奶粉以测定各鼠每日钙之存留量^[6],结果饲以全奶粉时钙之存留量平均数较饲以黄豆鸡蛋粉时多 1.91 毫克,标准差为 4.78 毫克,但经显著性测验结果并无显著差别,问若再进行同样实验时应将鼠增至几只?用上式即得

$$N = \left(\frac{1.96 \times 4.78}{1.91} \right)^2 = 24.1$$

故应将鼠增至 25 只。此系个别的比较时样本大小的估计方法,若为团体间的比较则用下式(7)(第三节中对这二种实验设计方法将进一步讨论):

$$N = 2 \left(\frac{ts}{x} \right)^2$$

$[s]$ 为二个样本合并之标准差, $[N]$ 系指每个样本的大小,余同前式。例如上述同一报告中用鼠两组进行,甲组 12 只,饲以黄豆鸡蛋粉;乙组 9 只,饲以全奶粉,均测定其每日氮之存留量。结果甲组所得平均数较乙组多 20.82 毫克,合并标准差为 28.06 毫克,经显著性测验后为不显著,但其结果中似有区别之趋势,不显著之产生可能因样本太小,用上式计算可得出二组之估计样本各应用 14 只鼠。算式为

$$N = 2 \left(\frac{1.96 \times 28.06}{20.82} \right)^2 = 13.9$$

当仅知全距而不知标准差时可利用二者之关系估计标准差之近似值。下表中 $[N]$ 系样本大小, $[R/s]$ 系全距与标准差比值之平均数。如过去文献所载调查我国二十一岁男子 226 人之体重^[8],其全距为

$$73.5 - 44.0 = 29.5 (\text{公斤})$$

下表中相当于 $N=200$ (因与 266 最接近) 之 $R/s=5.5$, 故

$$s=29.5/5.5=5.36(\text{公斤})$$

即为该年龄男子体重标准差之估计值, 从原始资料直接计算之标准差为 5.43 公斤, 可见二者极接近。

全距与标准差之比值(1)

N	R/s	N	R/s
2	1.13	20	3.73
3	1.69	30	4.09
4	2.06	50	4.5
5	2.33	75	4.81
6	2.53	100	5.02
7	2.7	150	5.3
8	2.85	200	5.5
9	2.97	300	5.8
10	3.08	500	6.1
15	3.47	700	6.3

(2) 计数资料中样本大小之估计——统计中另一类资料是计数的, 例如康氏反应阳性与阴性人数, 某病患者人数, 某种疗法治疗后痊愈、进步、无效及死亡人数, 药物试验中动物之死亡与生存头数, 自然界中感染疟原虫之蚊虫等都属于计数资料, 这类资料往往可以求出百分比(或千分比等), 其样本大小的估计用下式:

$$N = \left(\frac{t}{p_1 - p_2} \right)^2 [p_1(1 - p_1) + p_2(1 - p_2)]$$

$[p_1]$ 与 $[p_2]$ 为两个样本各自的百分比(或千分比等), $[N]$ 为每个样本的大小, 其他同前二式。例如某市 1953 年用血浆疗法试治流行性乙型脑炎患者 63 人, 病死率 20.6%, 同时用一般疗法治疗同样患者 58 人, 病死率为 32.8%, 但经显著性测验为不显著, 原文谓“根据统计分析, 不能证明血浆治疗在降低病死率上具有肯定的功效, 两组间之差别因材料太少缺乏充分的依据, 说明其并非由于机率使然, 所以血浆的疗效有待于与其他各地区作比较或收集更多的资料再作进一步的分析。”^[9] 如果问再作进一步的分析时至少应增加人数至若干为宜, 则可将此项初步结果代入上式:

$$N = \left(\frac{1.96}{0.328 - 0.206} \right)^2 [0.328(1 - 0.328) + 0.206(1 - 0.206)] = 99.1$$

结果表示应将二种疗法之病例均至少增至 100 人。

(3) 间杂性资料中样本大小之估计——一致性全体中样本越大可使误差越小,但在间杂性资料中则未必如此,此时应先行分组,然后再在各组中分别抽样。如果全体共有 35000 个单位,按其组成和变动可分为 3 组(见下表),问欲在此全体中抽取 1000 个单位的样本,各组应抽取多少? 计算方法如下表(1),结果表示第 1,2 和 3 组各应抽取 338,519 和 143 个单位,得到共 1000 个单位的典型抽样的样本。

间杂性资料中样本大小之估计

分组	各组次数 K	标准差 σ	$K\sigma$	样本大小 $n = K\sigma(N/S)$
1	10000	13	130000	338
2	20000	10	200000	519
3	5000	11	55000	143
合计	35000	—	$s = 385000$	$N = 1000$

各组确实之变异未知时可用样本之估计值,若仅知各组间之比例而不知标准差时则可使 n 与相当之 k 成正比,虽其精确度不若上表,但已尽可能地利用了仅有的知识了。

三、实验设计方法

实验设计应根据需要与可能而选取适当的方法。兹举出以下六种设计方法供作参考。

(1) 个别的比较——对同一个观察单位先后进行两次观察,而以两次记录的相差作为研究的结果。如在某种贫血症的疗效研究中可对治疗前与治疗后的血球计数作比较的观察,然后以其先后两次计数的相差求的平均值。另如试验某种药物短期间对蚊虫的杀灭能力亦可用各实验区施药先后的蚊虫密度之相差平均值作为衡量效果的数值(此时之观察单位是施药之各实验区)。这种对每一个单位先后作两次观察,而在第二次观察之前施以某种处理的实验设计,其优点是各单位在两次观察中其本身的各种条件大致相同,因此,一般说来其差数本身并未含有误差(但各差数之间仍然是有的),只要控制严密,记录正确完整,就可以得到比较可靠的结果。在选择观察单位时必须是同质的,

各单位与实验结果有关的各种条件如年龄,性别,体重,健康状况或病情之轻重等必须尽量使之一致。二次观察记录差数之平均值可作“ t 值”测验(其全体平均值为零)。

与个别比较类似的方法是配对,因为有些实验是不可能对同一单位进行二次观察的。例如临床治疗中,一般是不能把二种疗法先后施于同一病人的,要比较这二种疗法何者为优可将病人按病情之轻重及其他有关条件配对。又如某些被实验的动物在一次实验后即牺牲。亦可按照窝别,性别,原始体重等配成每二只为一对。配对后可用黑白各半之玻球一袋,随机取出一球若为黑色则第一对中第一个动物予以甲种处理,第二个动物则予以乙种处理;放回玻球,在袋内彻底搅匀后再随机取出一球,如为白球则第二对中第一头动物予以乙种处理,而第二头动物予以甲种处理,如此继续进行至最后一对。若处理方法在二种以上时则各组动物应同样增加,袋内玻球种数亦同样增加(各种颜色之玻球数仍应相等)。这样,每一对(或组)中各动物的有关条件都控制地相似了,只是受到不同的处理而得到各自的结果。所以,用二个(或二个以上)动物的配对观察也可看作是同一个动物先后二次(或多次)观察的设计一样,其统计处理方法亦同。

(2)团体的比较——有些研究是不宜对每个个体作先后观察亦不应配对的。例如要研究青春发育期间男性和女性某些生理常数的变化,这时我们可以控制的因素是年龄,性别以及一般的健康状况等条件,但我们没有理由把张姓男子和李姓女子作为观察的配对,这是没有根据的。又如某种急型传染病的预防接种试验中,只要分成接种与对照二组,通过一个感染的时期之后各有一个发病率而比较其发病率。在测验两种药物对小白鼠的半数致死量时则各小白鼠按所用药物分成二组,各组均将得出一个半数致死量,此时比较的内容是两个半数致死量,这些都是集团的结果。实验设计时可将小白鼠按体重排列,而以装黑白各半之玻球一袋随机摸取一枚,若为黑色则第一鼠予以甲种处理,放回袋内,搅混后再取第二球,若为白色则第二鼠予以乙种处理(注意此处理随机摸球之方法与前面所用者不同),如此一直进行至最后一鼠,结果两组的组数不一定会完全相等,但两者也不会相去太远,因为黑白球出现的机遇是相等的,只是通过随机取样的方法有些偶然的差别罢了,而这种二组数目并不完全相等的设计并不妨碍实验的结果。这种统计处理在测量资料中可用两个样本平均值相差的“ t 值”检验法,多于二组时玻球种类亦应同样增加,各种球数仍相等,结果用变异数分析法处理之;计数资料可用 χ^2 (Chi-square)测验。

统计上还有用各种随机数目的方法来使抽样随机化,随机数目表及其使用方法可参阅专著(1)(10)。