

教育与心理统计学

左任侠编著



华东师范大学出版社

教育与心理统计学

左任侠 编著

华东师范大学出版社

教育与心理统计学

左任侠编著

华东师范大学出版社出版

(上海市中山北路3663号)

新华书店上海发行所发行 沙洲印刷厂印刷

787×1092毫米 32开本 14 1/2印张 325千字

1982年2月第1版 1982年2月第1次印刷

印数1—10000本

书号：7135·039 定价：1.45元

前 言

《教育与心理统计学》乃60年代初期“向科学进军”的产物。当年编写此书的动机是旨在为我校教育和心理专业的学生以及中、青年教师从事科研时，掌握一些必备的统计学方法而提供的第一部讲稿，曾于1962年承蒙华东师大出版社刊印，并在校内发行。当时的着眼点就仅仅只把某些统计方法应用于有关教育、心理实验的设计和统计分析上面，而应用于教育、心理测量方面的那些统计方法，拟另写一部《教育与心理测验》，来完成这一整套教育、心理科研工具的任务。

未能逆料的是，后一部讲稿详细提纲于62年被人窃去，而那些有关内容的资料在“文化大革命”中又遭冲击，荡然无存。幸而全国心理测验专家们于1980年度的武汉会议后重振旗鼓，从新的角度通力协作，展开研究，行见老残未竟之志，旋将体现于这次会议成员的艰苦而又前途光明的奋战中。

华东师大心理系的领导同志们和同事们决定将老残60年代初期的旧本重新刊印，并正式公开发行，老残不胜惭愧，未能重新改写；幸得同事戴忠恒同志通过多次应用本教材于心理专业本科和研究生班进行实际教学的经验，不但把本书原印本的一些刊印错误加以校正，对原书少数段落亦略作变动；并且对本书索引重新作出修订。为此，老残特申谢忱。

最后，老残在此顺便表示一点遗愿：随着世界科技的进展，同时随着戴忠恒同志教学科研的进步，希望戴同志继续担负修订本书的责任，如果它能获再版的话。

左任侠

1981年9月于华东师大

目 录

前 言	(1)
第一章 绪 论	(1)
第一节 从科学谈起	(1)
第二节 实验、科学研究程序	(4)
第三节 统计方法与实验研究	(17)
第二章 次数分布	(22)
第一节 关于观察值的一般概念	(22)
第二节 次数分布表列法	(26)
第三节 次数分布图示法	(30)
第三章 集中量	(42)
第一节 关于集中量的一般概念	(42)
第二节 算术均数	(44)
第三节 中(位)数	(53)
第四节 众 数	(58)
第四章 差异量	(65)
第一节 关于差异量的一般概念	(65)
第二节 低效差异量	(66)
第三节 高效差异量	(74)
第四节 相对差异量	(82)
第五章 二项分布和常态分布	(87)
第一节 关于概率的一般概念	(87)
第二节 二项分布	(101)
第三节 常态分布	(108)

第六章 统计推理	(126)
第一节 关于统计推理的一般概念.....	(126)
第二节 抽样分布.....	(128)
第三节 全域参数的估计.....	(133)
第四节 统计假设的检验.....	(138)
第五节 一端检验和两端检验.....	(141)
第六节 在未知全域 σ^2 的条件下检验 全域参数 μ 的假设.....	(154)
第七节 检验有关全域比率的假设和有 限性乘数.....	(158)
第八节 施用 χ^2 量数作假设检验.....	(163)
第七章 相关和回归	(187)
第一节 相关量.....	(187)
第二节 回归线.....	(193)
第三节 相关系数与回归系数的归组计算.....	(205)
第四节 有关相关与回归的假设检验 及其抽样误差.....	(210)
第八章 比较两组研究结果的统计	(227)
第一节 差与和的变差.....	(227)
第二节 两个相关样组均数之差 的意义显著性检验.....	(230)
第三节 两个独立样组均数之差 的意义显著性检验.....	(234)
第四节 F 值和其它有关两个独立样组的 实验设计问题.....	(237)
第五节 比较两组的比率.....	(246)
第六节 比较两个相关量数.....	(251)

第九章 实验设计的逻辑	(258)
第一节 有关因果关系的推论.....	(258)
第二节 从实验研究中推出因果关系.....	(267)
第三节 从准实验研究中推出因果关系.....	(278)
第十章 简单随机化设计——变差分析初步	(290)
第一节 检验两个以上的样组均数的 全盘差异.....	(291)
第二节 初步变差分析的一般格式.....	(297)
第三节 全盘差异检验后的逐对差别检验.....	(304)
第四节 关于F检验的假定.....	(309)
第十一章 实验设计与变差分析续论	(327)
第一节 因子连乘设计.....	(327)
第二节 处理对层级设计.....	(354)
第三节 处理对受试设计.....	(364)
第四节 处理内分组设计和随机重复设计.....	(376)
第五节 拉丁方区设计.....	(386)
第六节 共差分析.....	(392)
附 表.....	(424)
名词索引.....	(448)

第一章 绪 论

第一节 从科学谈起

科学是遵循一定的研究程序所获得的规律性的知识体系。它的对象总是存在于现实世界各个领域内的、客观的事实——可以精确表述，可以分析、验证的事实。它的主要任务是在把客观的事实详为分类，明示其间的种种因果关系，从而抽出种种规律为它日实践提供具有普遍意义的理论基础。

让我们来先讲一个故事，然后根据这一故事来说明上面提出的有关科学性质的几个基本概念。

科学史上相传有一件公案发生在欧洲中古时期^①。一个女巫善治心脏病，凡往她家求治者只一喝她所配制的药剂，无不药到病祛。惟因当时的理发师，好象我国古时祝由科一样，都持有兼营医疗专业的执照；由于同行相嫉的缘故，她被同村理发师控以未获医疗执照、妖言惑众的罪名，卒至受到极刑的处分。在她未被控告之先，该理发师曾经收买一个心脏病患者暗中探悉女巫配制饮剂的秘密究竟是些什么。据云，这一神秘的饮剂是由以下数味药品调制而成：（1）团圆夜月之下所宰的蟾蜍血，（2）母兔的左后腿，（3）死马尾上的一根毛，（4）木桶盛取的天雨水和（5）一撮女巫后园生长的毛地黄花。女巫服刑后，理发师用同一秘方去治心脏病，果然也屡屡见效。不过，偶然一次在调制中错放了母兔的右前腿，发现其所煎出的饮剂为患者服后同样生

效；另一次，使用了月未圆宰杀的蟾蜍血去调制饮剂，亦同样奏效；又一次马尾毛被忘掉放进饮剂里，患者服后依然立愈；更一次因天旱长久未落雨，他使用河里的水煎药，结果亦能治愈疾病；最后，他省得麻烦，就直截了当地仅只采取女巫后园里的毛地黄花去医治心脏病患者了。实际上，现代医药界从毛地黄隔离出来的毛地黄甘（ $C_6H_8O_7$ ）才是治疗心脏病的药物。

这一由五味药材构成的秘方，更确切地称呼它，这一巫术虽说也能治病，可否便算为科学知识或医疗科学呢？尽管它尚能模糊地为实践服务，我们只能肯定它中间孕育着一点点科学的种子，但不是什么科学知识，更不是某种科学本身。这主要是因为此一巫术内面参杂着许多迷离恍惚、不知所云的东西。既未精确表述药剂与病理的联系，亦没有分析、验证每味药品与疗效间的关系。而且各味药品的配制方案更是不可思议。既说是母兔腿，何故专选后腿，而且还须左后腿？既说是蟾蜍血，何故要在月夜下，而且是圆月下宰杀的血？既说是马毛一根，何故要死马身上毛、而且是尾上毛？既说是水，何故要天雨水、而且是木桶盛取的水？既说是毛地黄花，何故一定要女巫后园的毛地黄花？仅就毛地黄花能够医治心脏病的关系而论，这一巫术——秘方——尚且不能把它阐明得足够清楚，那末，如果我们提高要求，要使这一秘方与心脏病的病理和药理结合起来加以表述，抽出某种规律、组织一种医疗体系，那更是不可能的事了。

然则那个理发师对于女巫的巫术——秘方一次又一次地“分析”、“验证”，最后居然达到“使用女巫后园的毛地黄花便可治愈心脏病”这一近于因果关系的认识，看来他所从事的工作岂不就是科学研究工作？他最后的认识岂不就是

科学知识吗？诚然，从人类科学发展的角度来看，他算是已经参加到科学工作的行列中来了，他的最终认识也包含着科学的内核了。但，他所进行的貌似实非的“分析”和“验证”全凭机会凑合来实现，而非依照一定的科学研究程序有意识、有计划地完成，因是我们就很难承认他是一位真正的科学家。又因他未曾有意识、有计划地遵循一定的科学研究程序进行工作，也就不可能揭露出真正的客观规律，更不能建立起一种理论体系，甚至不能把自己发现的结论纳入人类知识已有的体系内。所以我们也很难承认这就是科学知识。要知正确的、普遍的——当然是指相对的正确和普遍而言——科学结论须先经受实践的检验，才能在一定的理论体系内立得住足。在科学领域里，理论上站不住脚的东西，必定是实践中经不起检验的东西；实践中经得起检验的东西，必定是理论上站得住脚的东西。仔细审查一下理发师的结论吧。由于研究方法上的缺陷，他既没有真正揭露客观存在的规律，更没有建立任何具有普遍意义的理论体系，结果徒使他的认识局限于狭隘的经验主义范畴内，而经受不起实践的检验。科学的检验证明：不只女巫后园的毛地黄花可治心脏病，任何地方的毛地黄花都可产生同样疗效；不只任何毛地黄花可治心脏病，如果在一定的极微量配剂条件下使用药房出售的毛地黄甘（ $C_8H_{16}O_7$ ）亦可产生同样的疗效。②

近几年来，我国教育学和心理学工作者为了更进一步使教育学、心理学为四个现代化作出他们的贡献，正在努力建立教育学和心理学的科学体系，那么，我们对于那些从理论和实践中提出来的问题进行实验研究并切实地掌握有关实验、科学研究的程序，将不可避免地是向这一伟大目标迈出的第一步。

第二节 实验、科学研究程序

有人认为从事科学研究、实验（包括准实验）研究必有特定的方法、步骤，如同指导厨师烹调食物的“食谱”一样，能把每一样菜肴的烹调方法、步骤详详细细地开列出来，因而就希望指导科学研究的教师把每一项专题研究进程中的各个方面：如何开始、如何布置工作、如何解决难题、如何作出总结等等，一桩桩、一件件逐一开列出来。的确，这样做是办不到的。如果当真办到的话，这也不过是在研究某些早经解决了的、现成的问题。何况照这样培养出来的工作人员一定会缺乏独立工作能力呢！

但，实验、科学研究过程中究竟有否一般的程序值得科学工作者共同遵守呢？当然，这确实是有的。在从事论述这一共同遵守的程序之先，还有二事须得预先声明一下。第一，须明了实验工作中存在着两种不同性质的实验：一种是探索性实验；另一种是决断性实验。粗略地说来，探索性实验是从实际问题出发，当时尚没有找到一定的科学理论作背景的实验；决断性实验是在一定的科学理论上进行的实验。所谓在一定的理论上进行的实验，乃指在某一或某些假设的指引下进行的实验。这两种性质不同的实验就其进程而言，虽在起点上有所不同，然所应遍历的整个程序，固无不同之处。第二，马克思—列宁主义者认为真理是永不会完结的，所以每一次实验、科学研究遍历的程序只能视为长期追求真理过程中的一次循环。在承认了以上两点声明的前提下，且让我来谈一谈实验、科学研究的程序。这一程序大概可分为：（1）提出问题，（2）计划实验，（3）进行观察，（4）整理结果，（5）验证假设等五个环节。

一、提出问题——任何实验、科学研究总会是从生活实践中提出问题的要求出发，经过反复探索，结局乃能抵于问题的解决。可是，作为启程第一步的问题选择、提出，对于科学工作者，尤其对于初次从事实验研究的人的确成了一个问题。他们通常易犯的毛病有二。或者提出的问题太广泛，譬如说，“世界观与学习的关系”这类海阔天空、不着边际的题目。或者提出的问题太零碎，譬如说，“比较甲班学生与乙班学生的习字速度”这类囿于经验、昧于理论的题目。前一毛病的结果是：不仅得不出什么具体的结论，并且使得工作进程中在各个环节上都能遭遇到很大的困难。后一毛病的结果是：纵令得出某一具体结论，由于理论概括的困难，更由于结论在科学理论体系中孤立绝缘，使得它丧失了科学知识应具有的两个极有价值的特性，普遍性和连贯性。

针对以上两种情况，所以富有经验的科学工作者在问题提法上一方面考虑如何使问题范围尽量缩小，另一方面更考虑使问题的意义在自己或旁人的理论体系上生根。关于这种照顾双方的考虑过程，可举一实例借作参考③。从前有人拟用音节作学习实验的材料，首先考虑选定的问题是“意义影响学习的速度吗？”马上发现这样的提法不但研究的范围太嫌广泛，而且在实验设计、假设检验上都有困难。其次，考虑另一提法，即“具有不同意义的各种音节表现于学习时间上，材料意义与学习速度的关系究竟如何？”后又发觉这样的提法，其研究范围虽加以适当限制了，但其研究意义表现得好象对于现存的学习理论毫无所知的样儿。最后，决定采用“检验学习速度为学习材料的意义丰富程度的直接函数”。这一提法的决定才足够显示：它既把研究范围加以约束了，同时也把研究的实际意义与当前的学习理论联系起来。照

这样瞻顾把理论与实际打成一片的前景来提出研究问题的功夫，虽不能要求初事科学研究工作的人马上做到，但，只要经常朝这个方向做去，终究会有一天做得到的。

二、计划实验——大约六十年前有一位教育实验专家仿佛这样说过：假如允许我六小时完成一项实验的话，我一定支配五小时作计划工作，而只留一小时去进行观察或实验。三十年后一位实验心理学家以事实肯定了这一句话的正确性。他于二小时内进行了一项简单实验，可是事前却化了20小时去计划它④。从这两位实验工作者的言行里我们可能吸取的教训是，计划工作应该是整个实验进程中的重点工作，同时也是实验成功或失败的关键性工作。任何一项实验，如果计划恰当，有效的结论必然会合乎逻辑地从实验资料中涌现出来。如果计划得不好，结局就会要么得不出什么有效的结论来，要么把自己主观的判断生硬地强加在实验资料上面。

大家知道，实验的神髓，总括为一个专词表达出来，就是控制。控制有两种涵义。一种涵义是指实验者所观察的事实是真实的，而从中概括出来的结论是接近于真理的，所以人类对于某一领域的现象逐次观察、总结得愈深入愈周密，我们就说这一领域的研究愈接近于真理，或者说愈受有控制。这是从科学的历史发展角度来理解它的。控制的另一涵义是指消除那些能够扰乱实验者对实验因子（或自变量）的效果作正确观察的无关因子的意思。每一实验至少有一个实验因子，另外有若干受控制的因子；实验者的任务是有意识地支配实验因子，使之引起预期的变化，同时控制其余的因子，使之稳定不变。此地所谓稳定不变并非指数学上常数不变的意思，而是指虽变亦能经受得住统计分析的相对稳定之意。在实验研究，特别是教育心理实验研究中所谓控制，一般

是指后种涵义而言。

在实验计划，或称实验设计这一环节里，我们通常要考虑以下几个方面。

(一)关于受试者的选择——就研究对象的数量而言，在控制较疏的实验中所抽取的受试者的数目应当较多，在控制较严的实验中所抽取的数目可以减少。但选择受试者的主要问题不仅在于抽样的大小，还在于怎样抽法和怎样分组才能达到适当控制的目的。

抽样的方式可概分为三类。一为随机抽样，用粗浅的话来讲，即是通过拈阄、掷骰或机遇数目表等技术，使得研究对象全域里的各个成分同样都有被抽选之可能的方式。二为有意抽样，即专对实验者认为具有代表性的对象加以抽选的方式；这一方式一般含有主观的系统偏差，但在抽选仅由三、五个对象构成的小样组时，却具有不可抹煞的优点。例如，实验者在只被允许抽选三、五人作学生家长收入研究的情况下，与其施用随机抽样，倒不如抽选确具代表性质的工人家庭子弟为宜。三为联合抽样，亦名分层抽样。它是联合以上两种抽样特征的混合方式。譬如，作教育心理实验时，先把全体受试者有意识地分为上智、中等、落后三个层级，而后把每个层级内的受试者又按随机抽样的方式从事抽选，然后分派于实验组和控制组。一般讲来，在实验室的工作条件下，随机抽样较易进行；在班级教学中进行工作的自然条件下，联合抽样比较合适；有意抽样在教育心理实验中比较少用。

至于受试者分组的方式，一般有：(1)单组，(2)等组和(3)轮组三类。由于受试者的分组方式不同，实验的布置亦随之而异。单组实验是指把通过随机抽样选取的一

组受试者，加以两次实验处理（A和B），或者说，把一组受试者布置于两种实验条件（A和B）下的方式。这种方式最简便，在实验室内有可能尽量消除无关因子的情况下最宜采用，否则，容易产生常定误差。等组是指把实验组的受试者和控制组的受试者在能力（或其它特征）上施以相等化的意思。等组实验最宜消除扰乱实验结果的无关因子，但在如何使两组相等的测定上比较困难。为了免除这一困难，研究者每每倾向于采用轮组实验。所谓轮组实验是意指两个或两个以上的单组的联合。轮组实验的布置是：假定在两组受试者中施行二种实验处理，每种处理均重复一次，在甲组中二种实验处理（A和B）的次序为先A后B，重复时则为先B后A。在乙组中施行二种实验处理时，就得轮转以上的次序如下：先B后A，重复时则为先A后B。这类分组实验的方法不特可以避免实验过程中产生时间上的误差，更能消除教育心理实验中由于教师的能力差别、教时、包括学生的个别差异等等方面无关因子所造成的系统误差。

（二）关于实验处理的次序——刚才提起的轮组实验，不仅牵涉到分组的问题，亦且涉及了施行实验处理的次序问题。如果一组受试者，譬如说，其总数为4人，如对每人施以一种实验处理，则实验处理的数目为4，试以A、B、C和D标示之；更如果把处理重复4次，那末，A、B、C和D四种实验处理的排列就成了问题。对于一个实验设计者就要善于审辨，譬如说，下面格式（1a）在避免前后处理间产生系统偏差上远较格式（1b）为优这一类的情况。

表[1.1] 实验处理轮转顺序的格式

		实验重复顺序						实验重复顺序			
		1	2	3	4			1	2	3	4
格式 (1a)	1	A	C	B	D	格式 (1b)	1	A	B	C	D
	2	B	D	C	A		2	D	A	B	C
	3	C	A	D	B		3	C	D	A	B
	4	D	B	A	C		4	B	C	D	A

倘若扩大组数使之增达4倍即作为研究对象的受试者的数目等于16,从而重复实验四次,自然各种实验条件(处理)的排列次序也就更加复杂化了。这是拉丁方区设计讲授中的问题,暂不申述。

(三)关于实验处理的性质和实现处理的方式——实验处理在性质上的规定在实验设计中占有一个极其重要的地位,因为它决定着实验的敏感性,决定着结果解释的意义。至于实现处理的方式更应是随着实验处理的性质为转移。当然,实现处理的方式如果布置得不恰当,反转来亦会影响结果的解释,损毁实验的敏感性的。

实验设计的任务是一方面使得无关因子受到控制化,另一方面使得实验因子(或自变量)有明确的规定,从而达到实验因子的每一微小的变动都能在受试者的作业反应(或应变量)上敏感地反映出来,并经受得起有效的检验。往往发现一些实验在整个进程中确实投入了不少劳力,并且在某些部分工作上亦能表现出一定的成就,但在具有决定意义的实验处理上其性质、其实现方式俱没有好好地加以明确的规定,以致最终的结论颇难经受严格的统计检验,尤难期其发挥实

验研究应有的意义——理论上明辨真伪、实际上解决问题的意义了。因此，实验工作者在概念上必须先将实验处理的性质规定明白，同时使得在概念上规定好的实验处理能在实际工作中转化为具体的实验条件，并且要有一定的方法将受试者感受实验条件影响的结果明确地测量出来。

设例以明之。当我们检验识字教学中“依形归类”方法优于“故事带字”方法这一假设时，如果一方面不把同识字教学密切联系着的智力、阅读能力等作为共变因子加以适当的控制，另方面更不把识字这一实验因子的两个变值（依形归类和故事带字）好好地从另一有关因子（学习材料的份量多寡）隔离开来加以控制或当作另一实验因子进行处理，结果必然是研究工作者对心目中的假设无从施行统计的检验和因果的检验，同时对搜集得来的资料也就无从作出明确的结论。

（四）参照统计分析的要求，记录实验结果——教育和心理的实验结果，如同其它研究的实验结果一样，大多是经过数量化的资料。而这些数字资料一定要按照适当的格式记录下来，以备下一步统计分析作依据。如果事前把记录格式准备好，以后的工作自会事半功倍；否则，任意记录、不依统计分析所要求的适当格式进行，匪特事倍功半，往往会导导致应做的统计整理工作成为不可能。设例以明之。斯莱特 R. B. Sleight 以阅读中错误的数目作判断标准，研究过刻度盘的各种形状对阅读准确性的影响^⑤。我们可以利用他的实验资料来作说明。假使现以五种形状的盘式：水平刻度盘 A，圆窗刻度盘 B，圆形刻度盘 C，半圆刻度盘 D 和垂直刻度盘 E 标志着五种实验处理。往后，更以不同速度对 5 个受试者分别施行上述 5 种实验处理，其结果假若不参照以后统计