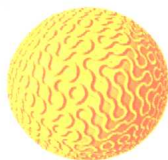


中 学 教 师 继 续 教 育 丛 书

JIAOYU TONGJI YU CELIANG JICHU HE JIAOYU PINGJIA

教育统计与测量基础 和教育评价

罗季重 编著



安徽大学出版社

中学教师继续教育丛书

教育统计与测量基础 和教育评价

罗季重 编著

安徽大学出版社

136894

教育统计与测量基础和教育评价

罗季重 编著

安徽大学出版社出版发行

(合肥市肥西路3号 邮码:230039)

肥西县印刷有限责任公司印刷 新华书店经销
开本 850×1168 1/32 印张 9.5 字数 238 千

1998年6月第1版 1998年6月第1次印刷

印数 2900册

责任编辑 李虹 封面设计 孟献辉

ISBN 7-81052-152-7/G·45 定价:19.00元

如有印装质量问题,请与出版社联系调换

内容简介

本书是中学教师继续教育培训教材之一。本书结合中学教师教学实际,介绍教育统计的一些基本特征量与基本方法:统计图表,集中量数,差异量数,标准分,均值估计与检验,相关与回归分析,方差分析, χ^2 检验;经典测验理论“四度”与现代测验理论;教育评价概念,评价信息处理方法,教学工作评价,教学评价量表与教师评价方案;数据处理微机程序,计算器在统计测量与中学数学教学中应用。本书特点是以介绍统计、测量、评价实用方法为主,不作过多理论推导和计算,数据处理均给出计算器和微机处理方法。

序

“八·五”期间,我院对初中教师继续教育做了一些有益的工作,取得了一定的成果。进入“九·五”,随着师资培训工作重点由学历补偿教育转到继续教育的轨道,在省市教育行政部门指导下,我院加大了继续教育的力度,加快了步伐。到2000年,要将合肥地区万名初中教师轮训一遍,并力争培养出500名教育教学骨干、100名学科带头人和若干教育教学专家。1997年9月,合肥市被国家教委批准为中小学教师继续教育实验区之一,作为初中教师继续教育的培训基地,我们感到责任更加重大,必须全力以赴地做好实验工作。

初中教师继续教育就目前的认识和理解,是指初中教师学历达标后,以提高政治业务素质和教育教学能力为目的的一种在职培训。初中教师取得相应学历,只能标志一个学习阶段的完成。随着社会进步、科技发展、知识更新,必须不断地接受新的理论、新的知识、新的技术和新的方法。参加在职培训是达到这一要求的途径。要培训,就必须有教材。出于这样的考虑,我们组织编写了这套供初中教师继续教育的丛书,既满足培训学员的要求,又希望能在继续教育的教学内容上作一些探讨。

这套丛书有以下几个特点:(1)创造性。继续教育和学历教育是两个截然不同的教育制度,使用两种截然不同的教材。前者侧重于对已达到某个学历层次,并具有一定教学实践经验的教师,进行更新观念、开阔视野、拓展知识、提高能力的教育;而后者,主要是为达到某个层次学历要求,而需具备的知识和能力。因此,继续教育的教材不是一个层次的知识照搬照抄,需要有全新的内容。

11249/09

(2)实践性。丛书编写的程序是：编写者深入不同类型的中学调查研究，了解初中教师在师德修养、学科前沿知识和基础知识、学科教育理论和教学能力等方面的需求，反复揣摩研讨，编写初稿；经过教学检验，保留受学员欢迎的内容，正式出书。可以说，丛书是从教学实践中来，再回到教学实践中去。(3)区域性。我国幅员辽阔，各地教师的构成和基础千差万别。丛书主要是针对合肥地区初中教师实际编写的，试图体现区域特色。对于开展首轮初中教师继续教育的地方来说，也有一定的参考价值。

初中教师继续教育是一项开创性的工作，其特色和规律，都处于摸索之中。因而，这套丛书也只能是一个阶段的教学成果。希望培训院校使用它时，及时和我们交换意见、提出建议，目的是加强教学内容的针对性、实用性、先进性和科学性，为提高初中教师继续教育的质量贡献一份力量。

丁元生

1998年5月

前 言

教育统计学与教育测量学是实现现代教育教学研究与管理的两门基础学科,是广大教育工作者进行科研教研必备的工具。而教育评价是教育科学研究重要领域之一,教育统计与教育测量为其提供搜集、整理、分析数据的基本方法。

“文革”后排除“左”的干扰,教育统计与测量在我国教育领域应用与发展较快,教育评价思想日益为人们所接受,并取得一定成绩。但由于我们在这方面的工作起步较晚,或由于历史的惯性,不少经验丰富的教师与教育工作者,对教育教学规律量化研究了解不多,或在教学研究和经验总结时还缺乏量化分析;因此有必要普及这方面基础知识。

90年代初以来,编者曾在合肥教育学院数学系数学专业师资格学历培训班、继续教育培训班和中学校长岗位培训班上讲授教育统计、教育测量与教育评价;不少同志希望有一本较为系统又深入浅出的实用教材,这促进我在原使用讲义基础上,经整理修改,编成此书。

本书主要特点:一是对理论公式不作严密推导,主要介绍实用方法;二是介绍实用公式时大多可以直接用计算器或微机处理数据。

本书第一章介绍教育统计、教育测量和教育评价概念;第二、三章主要介绍教育统计中描述统计和推断统计中一些基本方法;第四章主要介绍经典测验理论——“四度”,也对现代测验理论——题目反应理论作了简单介绍;第五章主要介绍教育评价有关概念,着重介绍了评价的信息处理方法,教学评价和学生质量评

价。本书在附录中给不熟悉概率统计的读者提供几个重要分布,并介绍数据处理的微机程序,计算器在统计与中学数学教学中应用,还给出了一些教学评价量表和教师评价方案。

本书可作为中学教师职务培训教材,也可作为中学校长提高培训教育评估课程的补充教材,以及中学教师自学统计测量与评价的教材。全书系统讲授需40-50学时;也可根据需要独立学习某些章节。如对非数学专业学员,可介绍第一、二章和第四、五章部分内容,而对有教育统计基础的学员,可直接学习第四、五章。

本书编写过程参考国内外有关书籍和教材,引用了一些资料,由于篇幅有限,不能一一列出出处。本书的附录二“数据处理的微机程序介绍”由合肥教育学院数学系钱泽强老师编写。在此一并表示谢意。

由于编者水平有限,经验不足和时间仓促,书中缺点与错误在所难免,恳请广大读者批评指正。

编者

1998年元月于合肥教院

目 录

第一章 绪论	(1)
§ 1.1 教育统计、测量与教育评价的概念及作用	(1)
§ 1.2 几个基本概念	(9)
第二章 数据搜集与基本特征量的统计	(14)
§ 2.1 数据搜集和统计图表	(14)
§ 2.2 集中量数和差异量数	(28)
§ 2.3 标准分	(36)
§ 2.4 直线相关	(47)
第三章 推断统计	(60)
§ 3.1 推断统计的基本原理	(60)
§ 3.2 总体均值的检验与估计	(63)
§ 3.3 两个总体均值差异的检验	(71)
§ 3.4 方差分析	(78)
§ 3.5 回归分析	(84)
§ 3.6 χ^2 检验	(94)
§ 3.7 抽样方法	(103)
第四章 测验的质量分析	(115)
§ 4.1 测验的性质、种类与功能	(115)
§ 4.2 测验的编制	(118)
§ 4.3 测验的信度与效度	(124)
§ 4.4 测题的质量分析	(139)
§ 4.5 题目反应理论	(147)

第五章	教育评价	(152)
§ 5.1	教育评价的概念与功能	(152)
§ 5.2	教育评价的类型与原则	(156)
§ 5.3	教育评价的程序与模型	(159)
§ 5.4	教育评价信息的处理	(168)
§ 5.5	教学工作评价	(179)
§ 5.6	中学生质量评价	(194)
附录一	概率与数理统计中的几个重要分布	(204)
附录二	数据处理的微机程序介绍	(210)
附录三	CASIO-570AV 计算器在教育统计中应用说明	(236)
附录四	TRULY105 计算器在中学数学教学应用说明	(243)
附录五	教学评价量表	(250)
附录六	中学教师继续教育调查问卷	(256)
附录七	中小学教师评价方案	(260)
附录八	附表	(269)
	表 1 标准正态分布密度函数数值表	
	表 2 标准正态分布函数数值表	
	表 3 χ^2 检验的临界值表	
	表 4 F 检验的临界值表(单侧检验)	
	表 5 F 检验的临界值表(双侧检验)	
	表 6 F_{\max} 的临界值	
	表 7 T 检验的临界值表	
	表 8 肯德尔 W 系数显著性临界值表	
主要参考书目		(295)

第一章 绪论

§ 1.1 教育统计、测量与教育评价的概念及作用

§ 1.1.1 教育统计

1. 统计及教育统计的产生与发展

说到“统计”这个名词,大家都不陌生,它是一种搜集数据方法。但这远非统计的全部内容,我们所说统计,是作为一门科学,即统计学,它是对于事物总体信息的搜集和分析,并以此为根据进行推断的理论和方法。我们学习的教育统计学是应用统计学的一个分支。它是探讨如何应用统计学的一般原理和方法,特别是数理统计方法,来研究教育(掌握教育情况,探索教育规律,制定教育方案,检查教育效率等一系列教育问题)的一门科学。

统计的起源很早,远在人类社会初期,就有对某种事情用打绳结、堆石块、画道道等方法进行简单记数,这便是统计的萌芽。可以说,有了计数,就有了统计。随着人类社会的发展,这种计数方法也随之发展、提高。在中国的夏商时代以及古埃及、古希腊、古罗马都有人口、居民财产等统计或人口调查制度。但这些都是统计实际活动历史,而统计作为一门科学,即统计学产生是近代的事。它起源于17世纪中叶德国的“国势学”与英国的“政治算术”。随着概率论的产生与发展,才使统计学理论真正发展起来。特别是比利时学者凯特勒(A. Quetelet)把概率论中大数定律、误差法则、正态分布等概念和计算方法引入了统计学。他运用这些理论

研究人口、犯罪及人体测量等问题。他首先在智力测验方面提出这样的思想：“人的道德品质可以用他的行为来度量，而智慧能力可以用工作（著作）来度量。”1835年，他试图发现写作与年龄关系，他搜集了关于165个剧本的47位法国作家资料进行考察，最后的结论是：剧本创作才能出现在21岁以后，30-35岁是精力旺盛时期，50-55岁是成熟时期，以后便日渐衰退。这可以说这是统计在教育方面应用的早期史例之一。

因为统计学在工业、农业、贸易、司法、卫生、教育等各方面应用，而逐步形成各专门统计，教育统计就是其中一支。

教育统计学作为一门独立的科学，直到本世纪初才建立。几十年来，随着人们对教育现象数量特征的认识，由精确性、随机性到模糊性的不断深化，教育统计也从以描述统计为主的阶段，发展到以概率分布理论为中心的推断统计阶段，进而又出现模糊统计方法。同时由于计算机科学的发展，给教育统计学添入了更大活力，发展得更快。

我国辛亥革命后，随着西方科技教育的发展，教育统计学也被介绍进来。解放前，教育统计学是高等学校教育系及师范院校的一门必修课，统计方法在教育科研和教育管理方面，曾得到一定程度应用。解放后，由于“全盘苏化”，教育统计学从院校课堂消失，也无人从事这方面研究，直到1978年以后，我国各师范类院校才陆续开设这门课，近二十年来，我国教育统计学发展很快，在教育科研和教育管理上应用也日益发展。

2. 描述统计、推断统计与模糊统计

描述统计 对于所关心的教育现象进行全面调查或观察，然后将所得到的大量数据加以整理，简缩制成统计表或统计图；或就这些数据分布特征，如集中趋势、离散趋势、相关程度等，计算一些具有高度概括性的特征数，如平均数、标准差、相关系数等；借助它们，我们可以从杂乱无章的资料中取得有意义的信息，便于对不同

总体进行比较,作出结论。在教育实践中,使用描述统计的机会很多,如各级教育行政部门在统计各级各类学校在校学生数,教师要了解同年级各班平均成绩及标准差,再如男、女生赛跑速度的差异,学生理化成绩与数学成绩是否相关等。

推断统计 在教育实践中,对许多现象不可能进行全面调查,或全面调查困难太大,这样得不到描述统计的数据,这时就可以进行推断统计。推断统计是以概率论与数理统计为其理论基础的,根据随机抽样的原则,抽取总体中部分个体组成一个整体(称为样本),通过对样本的观察,得到一些信息,然后用一定数理统计方法,进行分析,在一定可靠度上推测相应的总体特征。简言之,就是从部分推断全体。比如我们想了解全市5万名初中生患近视(视力在5.0以下)者所占比率,不同近视程度及各年级近视状况等。若假定由于时间和经济上原因不允许进行普查,则较好的方法就是从全市5万名初中生中随机抽取部分学生,比如1000名进行视力检查,然后根据所得到的各种数据推断总体的相应数据。若1000名学生中患近视者有250名,即占25%,则经过统计处理,我们就可以指出“能以95%的置信概率(即有95%的把握)推断全市初中生患近视的百分比为22.3%~27.7%”。在这个例子里,全市5万名初中生的视力构成一个总体,而抽出1000名学生的视力构成一个样本。像这样根据得到样本的数据来推测总体的性质,并标明可能发生误差的统计方法,就是推断统计。这类统计方法在现代教育统计应用的相当广泛。

模糊统计 模糊数学是研究和处理模糊信息的工具,这是60年代产生的新的数学分支。国内介绍这门科学,才有十几年的历史。模糊数学一产生,就为模糊统计奠定了方法论基础,使人们对事物数量特征认识,又多一类统计工具。教育领域模糊现象比比皆是,如学生品德优劣、解题能力高低、教师教学能力高低、教学效果好坏等,这些都无法用描述统计或推断统计作定量描述,只能作

定性的描述;有了模糊统计方法,就可以对它们进行定量分析。如对教师的课堂教学效果,可按评价要求分为教学目的、教学内容、教学过程、教师基本功、教学效果等五个因素;评价结果可分为很好、较好、一般、较差四个等级;根据专家意见或经验,确定各因素权数分别为0.2,0.2,0.2,0.1,0.3。然后通过多人对某教师某节课的评价,得到单因素评判矩阵,通过模糊数学方法,进行综合评判,得出较正确结论。这就是模糊统计的方法。

这类模糊统计方法在教育统计中引入时间不长,但其生命力很强,必将成为教育统计的一个主要内容,特别为教学评价提供了一个强有力的工具。

§ 1.1.2 教育测量与教育评价

1. 教育测量与教育评价产生与发展

根据一定规则,对事物进行量的测定,就是测量。教育是人类有目的、有计划、有组织地培养人的活动。教育活动也离不开测量。教育测量指运用测量手段对教育活动所作的量的测定。凡是需要且能够测量的与教育有关活动都在研究之列,如教育投入、教育过程各要素、教育结果等。

按照一定的标准对客观事物作的价值判断,就是评价,这是人类社会生活中的一项基本活动。为了判断自己当前或过去活动的价值,人们每时每刻都在作出评价;它存在于人的一切有目的的活动之中。如一个人到商店买某种商品,常常对商品的规格、颜色、质量进行评比,最后作出抉择,在这个活动中就包含着评价。教育是一种有目的活动,所以在一切教育活动中都含评价。大部分教育评价都必须以数据来说明问题,而搜集、整理、分析这些数据就是应用教育统计方法进行教育测量。测验是教育测量的主要手段之一。

(1) 教育测验的产生及美国测验运动

教育评价理论产生,究其历史根源来说,是源于对学生的学力测验。我国西周《礼记·学记》中记载当时定期学业考查制度:“古之教者,家有塾,党有庠,术有序,国有学。比年入学,中年考校。一年视离经辨志,三年视敬业乐群,五年视博习亲师,七年视论学取友,谓之小成;九年知类通达,强立而不返,谓之大成。”真正的教育测量始于我国隋末开始科举考试制度。公元606年隋炀帝开始了历史最早的考试选才制,当时测验方法是帖经、墨义、策问、诗赋等。这种方式沿续1300年之久。我国科学考试作为教育测量开端,已为世人公认。18世纪西方各国学校考试主要是口试。1702年英国剑桥大学首先以笔试代口试,开西方学校考试笔试之先河。但由于开始笔试以主观题为主,其客观性较差,为矫正这一弊端,力求考试客观化,于是测验方法随之出现。在教育界引起人们对测验问题极大关注的是著名的莱斯(J. M. Rice)拼字测验(1894年)。人们称莱斯是教育测量的创始人。

教育测验在美国形成一个运动是本世纪上半叶的事,它经历三个阶段。第一阶段(1904-1915)开拓期,以桑代克(R. L. Thorndike)《心理与社会测量导论》一书发表为开始。桑代克等人提出“凡是存在东西都有数量”,“凡是有数量东西都可测量”,这一时期研究重点是将测验客观化和标准化。第二阶段(1915-1930)兴盛期,教育测验已发展到分为学历测验、智力测验与人格测验三种不同性质的测验。种类繁多,应用广泛。第三阶段(1930-1940)批判期,测验运动开始走下坡路,尽管教育测验运动不断发展,人们逐渐认识到,教育测验尽管能使考试客观化、标准化,并能把人的能力换算成数字,但毕竟不能测出人的全部,即使像学历测验也不能测得学历的全部领域。在使用测验的过程中,人们逐渐认识到,我们不仅需要鉴别,更需要决策,从而导致了从教育测验向教育评价的过渡。

(2) 教育评价产生与发展

1929年,美国发生一次规模巨大的经济危机,罗斯福总统在1933年实行“经济社会化政策”,大批青年没有就业机会,只好重返学校。但美国中学在大学入学考试这根指挥棒作用下,其课程设置主要是为升大学服务,不适应整个社会和失业青年的需要,这样学生需求与课程设置之间发生了尖锐矛盾。在这种情况下,美国一些教育家成立PEA(进步主义教育联盟),一方面批判旧的教育,一方面自我反省寻找新出路。他们选了13所大学、30多所中学进行实验,经过八年(1933-1940)试验;他们边试验边评价,提出报告,美国教育史上称之“八年研究”,这项报告被称为“划时代的教育评价的宣言”。报告强调,评价对试验成功起了积极重大作用。报告指出,实验中学的学生在社会、政治、经济、文化等方面的信仰与态度,远比传统学校的学生进步与开化;在独创性活动、集体协作性活动方面,其表现优于传统学校;只是记忆一项成绩不如传统学校,但差异尚未达到统计上显著性水平。因此,可以证明:实验学校的新课程优于传统学校的旧课程。

教育评价概念始于“八年研究”,历史上曾有不少人下过定义,但至今认识尚未完全统一。瑞夫·泰勒(Ralph. Tayley)在1950年曾首创教育评价定义,认为评价就是“确定教育目标在实际上被理解到何种程度的过程”。1953年克龙巴赫(Cronbach)等人提出“评价是为进行决策提供信息的过程”。豪斯(House)、史克列文(Scriven)、斯塔夫列宾(Stufflebean)等人一致认为“评价是一种既有描述,又有判断的活动”,是一种对优缺点和价值的评估。

目前被广泛接受的评价定义是1981年美国评价标准化联合委员会提出的“对教育目标以及对它们的优缺点和价值评估的系统调查,为教育决策提供依据的过程”。因为这个定义概括教育评价三个重要组成部分:①评价的依据——教育目标;②评价的手段——系统调查,包括定性和定量地收集资料的一切方法;③评价的作用——对教育活动作出价值判断,为教育决策提供依据。

教育评价发展至今,可分为两个阶段:

第一阶段(1940-1957)教育评价理论和方法不断发展,但评价活动未普及。美国“八年研究”关于教育评价的理论和方法被人们接受后,人们普遍认为,教育评价是教育领域中的重要课题。但是,为了使评价能够达到预期目的,关键是要明确每种教育活动目标,首先美国著名心理学家、教育学教授布卢姆(B. S. Bloom)在1956年完成认知领域(知识及知识运用)教育目标的结构研究,他把认知目标分为知识、领会、运用、分析、综合、评价等六项分目标。但情感和动作技能领域教育目标直到六七十年代才分别完成。

第二阶段(1957-)教育评价理论体系逐步完善,教育评价活动逐步普及。这个时期教育评价由教育工作者所关心的课题逐步变成各级政府和广大民众所关心的课题。1957年苏联人造地球卫星上天,美国人分析自己落后的原因在于教育,教育落后震动美国政界和民众,于是美国中学教育遭到舆论的猛烈抨击。公众要求学校提出证明学校办学的成效。随后,政府进行大规模评价活动,评价由专家课题转入政府部门的教育措施。

1964年克拉斯沃尔(B. S. Krathwohl)与布卢姆完成情感领域目标分类:接受、反应、价值评价、组织、由价值或价值复合体形成的性格化等五方面。1972年哈罗(A. J. Harrow)与辛普森(E. J. Simpson)完成动作技能领域目标分类:反射动作、基本—基础动作、知觉能力、体能、技巧动作、有意沟通等六方面。这样教育目标体系基本形成,从而促进教育评价活动进一步发展。这段期间教育评价有如下三个特点:

(1)评价方法由早期注意“终结性评价”到注重“形成性评价”。

(2)充分强调自我评价作用,对评价结果采取慎重态度。

(3)新的科学方法被引入,注意定量评价与定性评价相结合,开展立体评价和全方位评价。

2. 我国的教育测量与教育评价