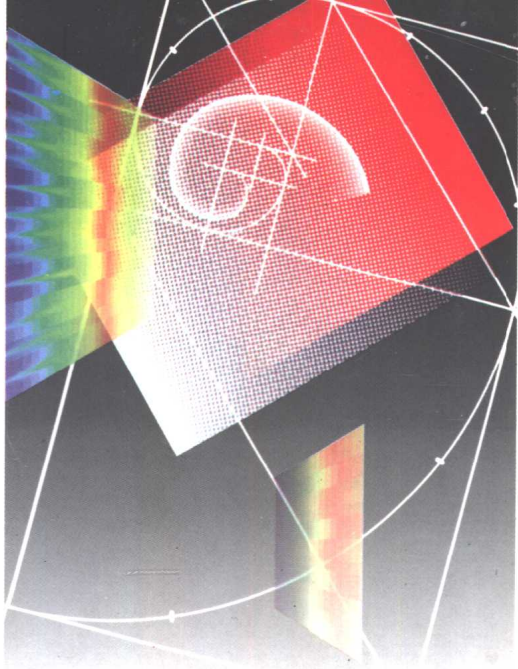




现

代

教育统计与 测评技术

朱德全 宋乃庆 主编

XIANDAI
JIAOYU
TONGJITONGJI
CEPING
JISHU

西南师范大学出版社

- 国家教委基础教育科研项目
- 西南师范大学出版基金资助项目
- 国家教委“九五”重点课题“西南贫困地区农村基础教育改革与发展综合问题研究”资助项目
- 荣获 1999 年第八届全国教育图书展优秀教育图书奖

现代教育统计与测评技术

朱德全 宋乃庆 主编

西南师范大学出版社

1998

责任编辑 李 红
特邀编辑 陈 红
封面设计 王 煤

现代教育统计与测评技术

朱德全 宋乃庆 主编

西南师范大学出版社出版、发行

(重庆 北碚)

新华书店经销

重庆电力印刷厂印刷

开本:850×1168 1/32 印张:16 插页:3 字数:407 千

1998 年 7 月第 1 版 1999 年 6 月第 4 次印刷

印数:23 501~29 500

ISBN 7-5621-1921-X/G · 1165

定价:19.80 元

序

张奠宙

人们做每一件事情都是有目的的,有目的的事情就得评价。一方面,评价行动的每一步是否朝着目标迈进;另一方面,当一件事情完成后,还要评价是否达到预定的目标。因此,毋庸置疑,教育需要评价,评价是教育工作的一个重要组成部分。要科学评价,就必须测量。因此,教育评价与教育测量紧密相连,教育测量是教育评价的直接基础,教育测量的结果是进行教育评价的重要依据,没有教育测量所提供的资料,教育评价是很难科学进行的。在进行教育测量时,会得到许多数据资料,这些数据资料必须进行统计处理与分析才能进行有效评价。因此,掌握教育统计学的基本原理和方法,是确保教育测量与教育评价客观科学的一个重要前提。

掌握教育统计学、教育测量学、教育评价学的基本原理和方法,长久以来就成为人们追求的一个重点。以往人们在学习和研究教育统计学、教育测量学、教育评价学时,往往将三者割裂开来进行研究,特别是开设这三门课程时,常常分期开设,不利于三门学科方法的相互运用,因而导致对理论的掌握不易深透,应用于实践时,往往走样,出现偏差。西南师范大学的朱德全、宋乃庆二位同志从解决这一问题出发,注重三者之间的内在逻辑联系,将三者联系起来进行研究,有利于节约教学时间,无疑也有利于读者更好地、更系统而又全面地掌握其基本原理和操作方法,这也是该书在此领域研究的一个新尝试。

《现代教育统计与测评技术》一书,在编写过程中注重汲取当今国内外教育统计学、教育测量学以及教育评价学的先进成果,同时更注重方法的介绍,注重提高读者综合运用各种方法的能力。通览全书不难发现该书具有以下特点:体系新——将三者紧密联系,有助于读者整体把握;实践性强——将理论与实践紧密联系,有助于读者灵活运用基本理论;反映及时——各种范例都力求体现现代性,有助于读者用现代先进的教育研究成果指导自己的实践;简明易读——文字简明扼要,分析深入浅出,有利于广大读者学习。相信本书的读者对教育统计与测评技术会有一个更深刻的领会,以指导自己的实践。

教育统计与测评技术是教育研究者们深感需要而又苦于很难系统学习掌握的一门学问。本书的出版将有助于初学者尽早入门,使其系统而又全面地掌握教育统计与教育测评技术。也有助于已有基础者更深入、更系统地把握教育统计与教育测评的基本原理和方法,使他们的教育研究水平更上一个台阶。同样,对开展教育统计与测评理论的研究,该书又提供了一种新思路,即将教育统计学、教育测量学与教育评价学融为一体,使三门学科取长补短,扫清了读者学习的障碍,也有助于读者把握三门学科的逻辑关系,充分体现了该书“易教、易学”的特点。因此,我相信该书一定能得到众多读者的喜爱。

1998年7月于
华东师范大学

(编者注:张奠宙系华东师范大学数学系教授,著名数学教育家,国际数学教育委员会执行委员)

前 言

本书将教育统计、教育测量与教育评价联系起来进行研究,这不仅符合这三门学科之间的逻辑关系,也有利于对这三门学科研究的取长补短。本书的编写目的是使读者能比较系统而又全面地掌握教育统计、教育测量以及教育评价的基本理论和操作方法。

本书是编者多年来为本、专科生和研究生授课讲义的基础上进一步新增、整理、充实、修改完成的,也是编者在教育统计、教育测量与教育评价的理论和实践探索中,根据多年来研究所取得的成果的基础上撰写而成的。

本书可作为师范院校(包括教育学院)本、专科及研究生(可分不同层次选取内容)教育统计、教育测量与评价的教材,亦可作为教师、教育科研工作者、教研员对教学质量、教改实验进行测评的参考用书。

本书的编写原则是:

1. 着力于新体系的建立。本书分三篇研究,即教育统计篇、教育测量篇以及教育评价篇。教育统计篇分三大块研究,即描述统计、推断统计和实验设计;教育测量篇包括测量的基本理论和方法,突出介绍良好测验的特征(信度、效度、区分度、难度等)以及编制标准化测验的步骤和方法,还特别介绍了现代教育测验中的一些实用量表;教育评价篇着力于研究什么是教育评价,为什么要进行教育评价,评价什么,用什么来评价以及怎样评价等内容。因此,本书力求体现思路的清晰性和广阔性。在内容编排上,由浅入深,

前后呼应,尽可能有效地帮助读者掌握现代教育统计与测评技术。

2. 着力于理论联系实际。本书注重方法的介绍,例证丰富,操作性强,有利于提高读者综合运用各种方法的能力。每一篇内容都是先介绍原理,再说明操作程序与步骤,最后以实例说明。竭力使读者既能把握有关原理,又能变通灵活运用有关原理。

3. 着力于适应现代教育的要求。本书汲取当今国内外教育统计学、教育测量学以及教育评价学的先进成果,适合于我国当前教育改革与发展的需要。书中各种范例都力求体现现代性,比如教育评价就是力求以素质教育为评价对象,强调评价个体的素质,强调评价学校对素质教育的落实情况等。

本书主要章节由朱德全、宋乃庆共同完成,其余部分由三位同事协作完成。张学敏:第十一章、第十七章;陈恩伦:第十三章、第十四章;杨爱民:第五章。全书由朱德全、宋乃庆统稿、定稿。

本书的出版得益于西南师范大学教科所所长张大均教授的帮助和指点,也得益于西南师范大学教科院领导的关心和支持,还得益于西南师范大学教科院研究生班学员们的好评,特别是苏飞跃同志的启发和鼓励。同时,我们还要特别感谢特邀编辑西南师范大学心理系陈红副教授的字斟句酌地审读全书,并感谢西南师范大学出版社编辑李红同志的精心加工。在本书出版之际,对以上给予我们帮助、鼓励、支持的同志表示真诚的谢意!

由于编者水平有限,错误、疏漏之处在所难免,敬请广大读者不吝赐教。

编者

1998年7月于
西南师范大学

目 录

教育统计篇

第一章 绪论	(1)
第一节 教育统计学概述	(1)
第二节 教育统计学的初步概念	(11)
第二章 描述统计	(16)
第一节 统计表与统计图	(16)
第二节 教育统计的特征量	(22)
第三节 利用电子计算器求特征量	(52)
第四节 回归分析	(54)
第五节 利用电子计算器求回归方程	(57)
第三章 推断统计	(60)
第一节 二项分布	(60)
第二节 正态分布	(64)
第三节 总体平均数的估计	(69)
第四节 几种常用统计检验方法	(73)
第五节 总体比率的推断	(85)
第六节 方差分析	(91)
第七节 聚类分析	(105)

第八节 非参数检验·····	(109)
----------------	-------

第四章 实验设计 ·····	(124)
-----------------------	-------

第一节 实验设计的操作要素·····	(124)
第二节 实验设计的基本模式·····	(126)
第三节 实验设计的主要内容·····	(136)

教育测量篇

第五章 教育测量概述 ·····	(157)
-------------------------	-------

第一节 教育测量的含义与特点·····	(157)
第二节 教育测量的发展史·····	(159)
第三节 教育测量的要素和种类·····	(165)
第四节 教育测量的功能·····	(171)
第五节 良好测验的特征·····	(173)

第六章 信度 ·····	(176)
---------------------	-------

第一节 信度的计算·····	(176)
第二节 影响测验信度的因素·····	(198)
第三节 提高测验信度的方法·····	(200)

第七章 效度 ·····	(207)
---------------------	-------

第一节 效度的估算·····	(207)
第二节 提高测验效度的方法·····	(220)

第八章 难度 ·····	(227)
---------------------	-------

第一节 难度的计算·····	(227)
----------------	-------

第二节	难度对测验的影响·····	(231)
第三节	难度的转换·····	(234)
第九章	区分度 ·····	(240)
第一节	区分度对测验的影响·····	(240)
第二节	区分度的计算·····	(243)
第十章	教育测验的编制与实施 ·····	(256)
第一节	编制测验的一般原则·····	(256)
第二节	编制测验的一般步骤·····	(258)
第三节	各类试题的编制·····	(264)
第四节	测验的编辑和实施·····	(269)
第十一章	各类教育测验 ·····	(274)
第一节	智力测验·····	(274)
第二节	学业成就测验·····	(303)
第三节	人格测验·····	(318)
第四节	品德测验·····	(328)

教育评价篇

第十二章	教育评价概述 ·····	(337)
第一节	教育评价的概念·····	(337)
第二节	教育评价的类型·····	(339)
第三节	教育评价的功能·····	(342)
第四节	教育评价的原则·····	(343)

第十三章	教育评价的指标体系	(349)
第一节	教育评价指标体系的要素	(349)
第二节	教育评价指标体系的准则	(354)
第三节	教育评价指标体系的设计	(356)
第十四章	教育评价的模式与程序	(361)
第一节	教育评价的模式	(361)
第二节	教育评价的组织	(365)
第三节	教育评价的程序	(369)
第十五章	教育评价的方法与工具	(374)
第一节	教育评价的方法	(374)
第二节	教育评价的工具	(383)
第十六章	教育评价心理与调控	(389)
第一节	教育评价心理与调控概述	(389)
第二节	主评者的心理与调控	(392)
第三节	被评价者的心理与调控	(401)
第十七章	教育评价的内容	(408)
第一节	学校评价	(408)
第二节	教师评价	(414)
第三节	学生评价	(425)
第四节	几种教育评价指标体系举样	(432)
统计用表		(448)

第一章 绪 论

第一节 教育统计学概述

一、教育统计学的研究对象

统计学是研究统计原理和统计方法的一门科学,它包括数理统计学和应用统计学两大分支。其中数理统计学主要是以概率论为基础,对统计原理和方法给予数学证明,对统计数据的数量关系模式加以科学解释,它是数学的一个分支。应用统计学是数理统计原理和方法在各个领域中的应用,它与研究对象紧密相关,如数理统计的原理和方法应用到工业领域,称为工业统计学;应用到医学领域,称为医学统计学;应用到教育领域,称为教育统计学。数理统计学与应用统计学联系很紧密,一方面,数理统计学是应用统计学的理论基础;另一方面,应用统计学是数理统计学的实践和应用,它使数理统计学的内容更加丰富和完善,也为数理统计学提出了实践中需要解决的新问题。

教育统计学是把数理统计学的原理及其方法应用于研究教育问题的一门应用科学。它是研究如何收集、整理、分析由教育调查和教育实验所获得的数据资料,并以此为依据,进行科学推断,揭示教育现象所蕴含的客观规律的一门科学。

教育统计学是教育科学研究的一个重要分支。它为教育研究

提供了一种科学方法,是教育科学研究定量分析的重要工具。它不仅有助于教育管理工作提高管理水平,同时也有助于教育理论研究者查阅有关教育研究的文献,更有助于教育实验设计的科学合理。

二、教育统计学的研究方法

教育统计学作为一门独立科学,不仅有它自己的研究对象,而且也有它自己的研究方法。要探究教育统计学的研究方法,首先必须明确教育统计工作的任务。教育统计学的全部研究过程包括密切联系的三个阶段,每一阶段各有其任务。为完成统计工作各阶段的任务,则有各种不同的研究方法。

(一) 教育统计学的研究过程包含三个阶段

1. 教育统计资料的收集 —— 第一阶段

教育统计工作的成败,与教育统计资料的收集密切相关,因此,为确保统计工作的顺利进行,必须慎重地收集原始资料,在收集资料的过程中必须遵循三条原则。

全面性原则,即统计资料必须全方位、多角度地收集,使统计资料具有代表性和典型性。

准确性原则,即统计资料必须力求科学性,尽可能减小误差,排除干扰,确保统计资料准确可靠,否则就有碍于后期工作的顺利进行。

客观性原则,即统计资料必须实事求是,不能弄虚作假,不能因主观需要而伪造数据,否则就违背了科学研究的精神。

2. 教育统计资料的整理 —— 第二阶段

教育统计工作第二阶段的内容包括用统计表和统计图对资料分组归类、检查核实原始数据、计算描述数据的特征量。教育统计资料经过这样整理,有利于下一阶段工作的顺利进行。

3. 教育统计资料的分析研究 —— 第三阶段

教育统计工作最后阶段的基本任务是对已经整理的统计资料进行分析研究。主要内容是对整理后的数据进行比较对照,分析新情况,发现新问题,揭示教育的规律和发展趋势,最后将分析研究资料的结果写成分析报告。

教育统计工作的三个阶段,虽说每一阶段的任务各有侧重,而且具有一定的独立性。但三个阶段却是一个统一的整体,它们相互关联、相互渗透,不能任意割裂。因此,任何一个阶段任务完成的好坏都将影响着整个统计工作的进程和质量。其中,收集统计资料是教育统计工作的前提条件和基础,整理统计资料是统计工作的关键,分析研究统计资料是统计工作的最终目的。所以,在教育统计工作中,我们应注重收集资料的准确性、全面性与客观性;整理资料的直观性和有效性;分析资料的科学性与系统性。只有这样,教育统计工作的质量才能得到保证,教育统计工作才能有效而顺利地进行。

(二) 教育统计资料的来源

教育统计资料的来源极其广泛。概括起来,主要有四大来源,即观察、调查、实验以及文献资料,其中主要资料来自于教育调查和教育实验。

1. 教育调查

教育调查是取得客观的、真实的可靠资料的一条重要途径。教育调查的方法和种类较多,就调查范围分,有全面调查和非全面调查。

全面调查即对全部对象都加以调查,如对某地区学龄儿童身体素质的调查,对某地区的中学生心理健康的全部调查等,都是全面调查。全面调查有自身的优点,也有其不足之处。优点在于可以全面了解某种情况,但不足之处在于费时、费力。

非全面调查即从调查总体中抽选具有代表性的一部分个体进行调查,因此,非全面调查可以节省人力、财力和时间。非全面调查

包括重点调查、典型调查和抽样调查,其中最常用的是抽样调查。

重点调查的调查对象在研究总体中所占比重较大,或比重虽不大,但在被研究现象的发展中起着重大作用。一般说来,按照调查任务的要求,凡在部分单位或少数地区能够反映所研究的项目和指标时,就可采用重点调查。在抽样调查以前,为了大致了解被调查单位的一般情况可以进行重点调查,在抽样调查以后,为了深入了解某一单位的情况也可以进行重点调查。可以说,重点调查是抽样调查的补充方法,二者可以结合运用。

典型调查是研究者根据需要选择一部分具有代表性的单位(或个体)进行的调查。典型调查可在较短时间内,用较少的人力、财力和物力,取得较大的效果。这种调查便于推广先进经验,树立先进人物。正因为如此,教育管理工作大都愿意采用这种调查方法。要搞好典型调查,关键在于选好典型。要选好典型,必须将调查对象按照某种标准分类,然后从每一类型中选出具有代表性的典型个体(或单位)。

抽样调查是根据概率论从总体的全部单位(或全部个体)中随机抽取一部分进行调查,并根据调查结果推断(或说明)总体的特征或规律。抽样调查可以以较少的人力、财力和时间进行更全面、更深入的调查。抽样调查必须遵循随机的科学原则,保证被研究总体中的所有个体都有同样被抽取的机会。所以,抽样调查又可称为随机抽样调查。

抽样调查分为单纯随机抽样、机械抽样、分层抽样和整群抽样。

(1) 单纯随机抽样

单纯随机抽样是从调查总体中完全随机地抽取调查单位(或个体)的一种方法。单纯随机抽样必须保证总体中每个个体被抽到的机会是均等的(即抽样的随机性),并且保证在抽取一个个体之后总体内成分不变(即抽样的独立性)。

单纯随机抽样可通过抽签以及随机数目表来实现。抽签法是先将总体中每一个个体都编上号码,再将每个号码写在签上,将签充分混合后,从中抽取几个(样本的容量)签,与被抽到的签号相应的个体就进入样本。随机数目表法是根据随机数目表随机抽取样本的方法。随机数目表是按随机抽样的原则编制的,每2个数字为1小组,每个区组包括25个小组,即每行5个小组,每列5个小组。根据需要,可将每个区组当作任意位数使用,从任何一项任何一个数目开始,从左到右或从上到下均可。比如,我们要从2000个学生中抽取100个学生作为样本,可先将2000个学生编号,然后从随机数目表中任意一个数字开始向任何一个方向摘录数字,以4个数字为1组,共取100组。假如我们从随机数目表的第11行第1列开始向右摘录数字:1818,0792,4644,1716,5809,7983,……所取的这些4位数中,凡大于2000直至4000者均减2000,大于4000直至6000者,均减4000,……使每一组数字都不大于2000,即1818,792,644,1716,1809,1983,……被编为这些号码的学生,就组成我们所需要的单纯随机样本。

(2) 机械抽样

机械抽样是按一个与研究问题没有直接关系的标志把总体各单位(或个体)加以排列,依一定距离机械地抽取被调查的对象。机械抽样可以通过把总体中的所有个体按一定顺序编号,然后依固定的间隔取样。例如,为了了解某校大学毕业班学生的英语学习情况,设全校毕业班学生总体有1500人,要抽取150人作为样本。可先将这1500人的英语测验分数由低到高排列,并从1至1500编号,然后按1,11,21,……号码的顺序和间隔抽取样本。

机械抽样比单纯随机抽样更能保证抽到的个体在总体中的分布比较均匀,而单纯随机抽样比机械抽样的随机性强,因为单纯随机抽样可以由总体内所有的个体组成任何可能的样本,而机械抽样对上例来说,只有10个可能样本:即由第1,11,……号分数组成

的样本,由第 2,12,⋯ 号分数组成的样本,⋯ 直至最后第 10,20,⋯ 号分数组成的样本。至于从中取哪一个作为样本可以随机确定。

机械抽样和单纯随机抽样二者也可以结合使用。如上例先按机械抽样原则在第 1,2,⋯,10 号分数中抽取 1 个,同样在第 11,12,⋯,20 号分数中也随机抽取 1 个,如此抽下去,所组成的样本既可保持分布的均匀,又扩大了各个个体随机组合的可能性。

(3) 分层抽样

分层抽样是把总体中各个个体按照一定标志分为不同类型或层次,然后从各类型中随机抽取若干个体,从而构成样本。将总体按照一定标志分类的基本原则是,层内之间差异要小,层外之间差异要大。否则就失去分层的意义。

例如,对某校 800 个学生的品德情况进行了解,拟取 40 个学生作为样本,即抽样比率为 $\frac{40}{800} = \frac{1}{20}$ 。那么,可先根据一定标准将 800 个学生分成优(160 人)、良(320 人)、中(240 人)、差(80 人)四部分,然后从各部分中用单纯随机抽样或机械抽样的方法,各抽取 $\frac{1}{20}$,即从优等中抽取 $160 \times \frac{1}{20} = 8$ (人),从良等中抽取 $320 \times \frac{1}{20} = 16$ (人),从中等中抽取 $240 \times \frac{1}{20} = 12$ (人),从差等中抽取 $80 \times \frac{1}{20} = 4$ (人),组成一个样本。

(4) 整群抽样

整群抽样是抽取的对象以整群为单位而不是以个体作为单位的抽样方法。这种方法的优点是便于组织,其缺点是代表性较差,因而采用也较少。

2. 教育实验

教育实验是指在预定的控制因子影响下,对教育方面有关的客观事实所进行的观察和分析。