

北京市高等教育自学考试委员会组编

教育评价与统计

周 谦 主 编

李国岚 副主编

科学出版社

1998

内 容 简 介

本书包括教育评价和教育测量与统计的基本知识。本书以现代教育学、心理学与数理统计为基础，运用统计工具和各种教育与心理测量手段，对教育效果评价的原理和方法进行了全面阐述，对测试项目和测验结果作了定性和定量分析，内容充实，结构严谨，书中有近百个实例，有利于读者学习理解，并紧密结合中小学教育和研究实践，对教育实践有指导意义。

本书可做高等教育小学专业课的教材，也可供广大教育工作者和高等师范院校学生参考。

北京市高等教育自学考试委员会组编

教育评价与统计

周 谦 主 编

李国岚 副主编

责任编辑 吕 虹

科 学 出 版 社 出 版

北京东黄城根北街16号

邮政编码：100717

北 京 双 青 印 刷 厂 印 刷

新华书店北京发行所发行 各地新华书店经售

*

1997年1月第一版 开本：850×1168 1/32

1998年8月第二次印刷 印张：13

印数：18 031—20 030 字数：340 000

ISBN 7-03-005847-X/G·704

定 价：19.50 元

(如有印装质量问题，我社负责调换〈环伟〉)

前 言

教育评价与统计，是心理科学、教育科学和统计科学的一个交叉学科。它是以现代教育学、心理学和教育与心理统计学为基础，研究运用各种教育与心理测量手段和统计工具，探讨对教育与教学效果进行评价的原理和方法。它既具有较现代的教育测量、统计和评价的理论知识内容，又有较强的实践性和可操作性。它是一门综合性很强的学科，也是实现教育评价、教育科学研究与教育管理科学化和现代化所必须借助的一门科学。

本书是高等教育自学考试小学教育专业学科体系中教育评价与统计课程的教材，试图使读者通过学习，着重理解教育测量和评价的理论与应用，掌握必要的统计方法和手段，并学会运用这些知识，对学校教育、教学和研究实践中的实际问题，作出比较正确与合理的判断和评价，以促进教育管理现代化、教育研究科学化。本书体系结构完整系统，既有较强的理论性，又密切联系中小学教育改革与评价的实际，具有重要的应用价值。资料翔实，文字简明易懂，便于自学，对中小学教师的教育科学研究与教育评价具有指导和参考意义。本书还可供广大中小学教师在继续教育中使用。

本书由北京高等教育自学考试委员会委托和批准，特邀请了首都师范大学周谦教授担任主编，北京教育学院李国岚副教授担任副主编。各章编写分工是：第一、二、三章由北京教育学院朱嘉泰副教授执笔；第四、五、十二章由北京教育学院李春山副教授执笔；第六、八、十一章由周谦执笔；第七、九、十、十三、十七章由李国岚执笔；第十四、十五、十六章由北京教育学院崇文分院童长江副教授执笔。最后请北京大学陈仲庚教授审阅了书稿。

本书是针对自学考试90学时的需要而编写的，每章后都配有

少量练习和思考题，供学习者练习使用。关于数据统计运算中需用的电子计算器，其型号和品牌种类很多，在本书中只能介绍几种通用的主要机型，供学习者参考使用。

本书编写过程中，参考了许多国内外有关教育评价、教育测量和教育统计的书籍和教材，借鉴了各书的编写经验，吸取了某些有益的观点并引用了一些可贵的资料和统计图表，特在此向诸位作者和出版者表示深切的谢意。

由于我们水平有限，加之时间仓促，书中缺点、错误在所难免，诚恳欢迎读者批评指正。

编著者

1996.7

目 录

绪 论	1
-----	---

第一篇 教育统计

第一章 数据的初步整理	7
§ 1 数据及其有关概念	7
§ 2 简单的统计图表	11
§ 3 次数分布表和次数分布图	15
第二章 集中量数	23
§ 1 算术平均数	23
§ 2 加权平均数	27
§ 3 中数和众数	30
第三章 差异量数	37
§ 1 百分位差	38
§ 2 方差和标准差	43
§ 3 标准差的应用	49
第四章 相关和直线回归	55
§ 1 相关的概念	55
§ 2 积差相关	60
§ 3 等级相关	64
§ 4 点二列相关	68
§ 5 直线回归	70
第五章 总体平均数的估计	77
§ 1 概率和正态分布	77
§ 2 总体平均数的估计	90
第六章 差异假设检验	99
§ 1 差异假设检验概述	99
§ 2 平均数差异假设检验	112
§ 3 相关系数差异假设检验	132

第七章 方差分析	141
§ 1 方差分析的基本原理.....	141
§ 2 单因素方差分析的步骤和举例.....	145
§ 3 各平均数间的比较.....	152
第八章 χ^2 检验	157
§ 1 χ^2 检验的基本概念与基本公式	157
§ 2 χ^2 的一般特性及其与 χ^2 检验的关系	159
§ 3 χ^2 检验的功能及在教育评价中的应用	162

第二篇 教育评价

第九章 教育评价概述	179
§ 1 教育评价的概念和功能.....	179
§ 2 教育评价的历史和发展.....	185
§ 3 教育评价的基本原则.....	191
§ 4 教育评价的一般过程.....	196
§ 5 教育评价的类型.....	201
第十章 教育目标和评价指标体系的建立	206
§ 1 教育目标的意义和分类.....	206
§ 2 小学教学认知目标分类表述和举例.....	210
§ 3 教育评价指标体系的建立.....	215
第十一章 有效测量与评价	224
§ 1 测试项目的难度.....	225
§ 2 测试项目的区分度.....	231
§ 3 教育测量与评价的信度.....	241
§ 4 教育测量与评价的效度.....	252
第十二章 评价学生的几种测量工具	267
§ 1 教育测量的特性.....	267
§ 2 学生学习成绩的测验.....	269
§ 3 智力与能力测验.....	279
§ 4 个性测验.....	287
第十三章 教育测量与评价结果的解释	291
§ 1 目标参照标准测验分数的解释和评价.....	291

§ 2 常模参照标准测验分数的转换和评价	296
§ 3 分数的组合和评价	304
§ 4 评价报告举例和交流	308
第十四章 学校管理工作评价	312
§ 1 学校管理工作评价的目的与意义	312
§ 2 学校管理工作评价的指标及内容	315
§ 3 学校管理工作评价的实施	321
第十五章 教师工作评价	331
§ 1 教师工作评价的涵义与作用	331
§ 2 教师工作评价的指标体系	334
§ 3 教师工作评价的组织与实施	346
§ 4 班主任工作评价	352
第十六章 课堂教学评价	356
§ 1 课堂教学评价的涵义与作用	356
§ 2 课堂教学评价的指标体系	359
§ 3 课堂教学评价的实施	369
第十七章 教育评价中的心理与调控	377
§ 1 教育评价中心理调控的意义	377
§ 2 主评人员的心理与调控	378
§ 3 被评人员的心理与调控	383
练习答案	391
附表	393
附表一 正态分布曲线下的面积	393
附表二 t 值表	395
附表三 积差相关系数 (r) 显著性临界值	397
附表四 相关系数 r 的 Z_r 转换表	398
附表五 等级相关系数 (r_R) 的临界值	399
附表六 肯德尔和谐系数显著性临界值	400
附表七 F_{\max} 的临界值 (哈特莱方差齐性检验)	401
附表八 F 临界值表	402
附表九 χ^2 值表	404
后记	407

绪 论

面临 21 世纪新技术革命的挑战,世界各国都认识到科学技术有力地推动着经济和社会的不断发展,同时,也清醒地看到科学技术又极大地依赖于全民族教育的不断提高,因此,几乎都把发展教育摆到了战略地位的高度.除了增加教育投资外,一个重要措施就是运用教育评价手段去实现教育目标.不少国家陆续建立了教育评价制度.我国政府早在“七五”规划中就指出:“加强教育立法,逐步建立系统的教育评价和监督制度.”又在 1993 年《中国教育改革和发展纲要》中再次强调“建立各级各类教育质量标准 and 评价指标体系”.教育评价已成为一个波及全球的潮流,影响着教育事业的发展.

教育评价与统计,作为教育现代化管理与指导的手段,发展得很快.它的理论与应用,已由对学生学习评价,发展到学校办学水平和管理评价,教师工作质量评价,乃至教材评价等.作为一名教师,要想了解怎样评价自己的教学工作,怎样评价所教学生的学业知识和技能,怎样使教学工作中考试环节更科学,就需要学习应用教育测量与统计工具,研究教育评价理论和方法.随着教育教学改革和研究的不断深入,如何测量和评价教育改革的成功与否,如何对教育质量作出科学评价,也需要研究教育评价与统计的理论和方法,因此,系统地学习和研究教育评价与统计的任务,就迫切地摆在了我国广大教育工作者的面前.

一、教育评价与统计学习的主要内容

教育评价是要对事物进行价值分析和判断,这就离不开测量.测量和评价两者密切联系着,又常称教育测量和评价.测量是评价的基础,是评价的重要依据,它着眼于客观数据的测定.教育

评价正是对测量所得数据进行判断的结果。因此学习和研究教育评价就必须学习和研究教育测量的理论和方法。这里我们所研究和学习的教育评价,就是包括教育测量与评价在内的理论和方法。

教育测量是依一定法则,用数字对人心理的一些特质行为加以确定,如对学生的品德、学习及其他能力、智力、人格等进行测量。对测量和收集到的数字,又叫数据,必须进行整理、统计和计算。根据对这些数据资料进行科学推断,才能做出评价。因此,教育评价既包含教育测量,又离不开教育统计学的原理和方法。一般来说,教育评价应该在定量基础上定性,以做出价值判断和决策,这就需要教育统计。所以说教育统计对教育评价来说是不可少的,在这个意义上说,没有教育统计就难以进行科学的教育评价。学习教育评价就需要首先学习教育统计学基本知识。为了便于读者系统学习,本书特在第一编首先较系统地阐述了基本的教育统计知识;第二编讲述教育评价的理论知识和方法。

(一) 教育统计学

教育统计是应用数理统计方法对教育领域中的问题,包括对了解教育情况、探索教育规律、制定教育方案和评价教育效果等进行研究的一门学科。它是应用统计学的一个分支,又是教育科学体系中的一个方法论的分支。教育评价、教育实验研究中所收集的资料和数据,都需借助统计学方法进行分析,才能得出科学结论。它是教育科学研究和教育评价中必不可少的工具。

教育统计研究的内容不断发展,大体分为描述统计和推断统计两部分。就发展顺序,描述统计居先,推断统计在后。

1. 描述统计。主要研究如何对所获得的数据进行整理、简缩和概括,进而描述数据的全貌。具体内容是,按一定方式整理数据;制作统计表和统计图;表示数据的分组和分布情况;计算能代表数据的某一特征量数,如集中量数、差异量数、相关系数等。

2. 推断统计。是指在描述统计的基础上,主要研究如何利用局部数据所提供的信息,如平均数、标准差、相关系数等,对全局的情况做出推断或判断。如调查全市10万名高中学生患近视眼

的百分比,而假设由于经济或时间的原因不允许进行全局普查,则最好的办法就是取其有代表性的一部分,如取1000人进行检查,然后根据这局部数据,经统计推断全市高中生患近视眼的99%的置信度的范围。又如,在教育研究中,从部分男、女学生阅读测验资料出发,然后研究推断在总体上学生的阅读能力有无性别差异。这都属于推断统计,在教育研究中是很重要的。本书在第一至第四章讲描述统计,第五至第八章讲推断统计。

(二) 教育评价

教育评价是把教育目标的一些原理和教育统计的一些方法,结合学校教育、教学,运用于对测验的项目和测验结果的分析,力图做出科学判断和决策的一门学科。它主要研究教育测量和评价的概念、功能、原则和过程;教育目标和评价指标体系;教育测量工具的编制、选择和质量分析;测量结果的解释和评价等。

1. 教育评价概述。主要讲述教育评价概念、功能、历史发展渊源、评价原则和一般过程等。见本书第九章。

2. 教育目标与评价。主要讲教育目标分类和表述,评价指标体系的研制等。见本书第十章。

3. 教育测量的工具和评价。主要阐述有效教育测量的难度、区分度、信度和效度分析,评价学生的几种工具,评价结果的解释。见本书第十一至第十三章。

4. 小学教育工作评价。主要介绍小学教育管理工作评价,教师教育工作质量评价,教师课堂教学评价。见本书第十四至第十六章。

5. 教育评价的心理和调控。见本书第十七章。

二、为什么要学习教育评价与统计

1. 学习教育统计,提高教育研究、教育评价水平。教育统计为教育研究提供了一种科学方法和工具,教育工作者掌握它就能应用它科学分析资料,处理数据,组织进行调查和实验研究,为教育评价、教育改革等工作提供客观科学的依据。这样就可以

使我们的教育研究和评价水平得到提高。

2. 把握教育目标，调整教育措施，跟上时代要求。学习教育评价，更自觉地看到当今世界新技术革命在科学技术和思想意识形态两方面向我们提出了严重挑战，迫使我们教育工作者坚持正确的政治方面，通过评价，不失时机地提出问题，把握教育目标，随时调整教育措施，跟上时代步伐。

3. 教育评价是教师工作的一部分，学习应用它促进并保证教学质量的提高。学业成绩测量和评价是教师教学过程中有机组成部分，对大多数教师来说既熟悉又陌生。说熟悉，是古往今来受教育的人都经过这个关口。说陌生，是不少人停留在旧经验的做法和观念上。教育测量的科学化包含的标准化测验，智力测量量表法等，国际上已研究实践了几十年，近年也受到我国教育工作者的重视。当前我们教育测评中也有很多不够科学的做法，自然对教育产生消极后果。例如，中小学考试中混淆了目标考试与常模考试的本质，一谈考试就排名次，导致学生片面追求高分，增加学生心理负担，也使人际关系恶化，产生了反教育的后果。又如，一些教育行政部门和学校过分重视总结性考试，“算总帐”，一旦发现问题就晚了。学习教育评价，帮助教师、学校选择适宜的测量工具，研究学生，研究教育，如分析学生学习能力，知识掌握程度，智力发展水平，以及时对教育教学过程和效果做出客观评价，及时调整教学，最大限度保证每个学生的发展，促进教学质量的提高。

4. 有利于教育决策科学化。运用教育评价与统计，决策编班、决策学生升留级，决策学生职业定向，决策学校教学管理工作优劣，决策教师教学工作质量和培训工作的方向，促进教育管理和决策的科学化。

5. 有助于学习先进经验和进行研究的交流。懂得教育评价与统计，使我们能顺利阅读国内外教育研究、教育评价报告和文献资料，学习他们的先进经验，进行学术交流，相互促进。教育研究和评价报告中常用一些统计方法和术语说明解释其成果。例如，

常碰到一些专门术语，诸如 $P < 0.01$ ；对虚无假设 H_0 加以拒绝； χ^2 值和检验； Z 分数；测量工具的信度 r ，测验项目的难度 P 等，都是一些常用的共同的术语和语言，如果不懂，就会妨碍吸收别人的研究经验。此外，自己的研究和评价也只有用这些教育统计和评价的术语说明，才能顺利地与同行进行交流。

6. 有助于提高科学思维能力，养成科学态度和科学精神。应用教育统计方法思考科学研究问题、教育评价问题，是一种思维能力、科学方法的训练，这对教育工作者是非常必要的。

三、教师应具备的教育评价与统计的知识和能力

教育评价是教育过程中必要的组成部分，作为一名教师，是教育过程的主导者，应具备起码的、从事教育工作所必须具备的评价与统计的知识和能力。

1. 了解教育评价的涵义，理解它在教育与教学过程中的重要意义和使用原则。

2. 理解教育、教学目标，能清楚、具体地表述学科教学目标，不仅懂得如何通过教育、教学达到目标，还能懂得如何评价达到目标的程度。

3. 能在教学中科学合理使用测验，在教育中选择使用测量工具，明确各测量工具的用途和局限。掌握试题分析和试卷可靠性与正确性的分析方法。

4. 能运用教育统计方法和工具，统计、分析、解释测量的资料和数据，对其结果给以评价，并用以指导教育评价，指导教育科学研究。

5. 能用现代教育观念，明确什么是一节好课，明确什么是优秀教师的工作质量，优良的学校教育管理工作质量，用以指导教师工作和参加教育评价活动，促进教育评价客观化和科学化，改进和提高教育工作。

教师还应有高尚的人格。教育评价工作中，要有高尚的道德规范，公正而合理地对待每位学生和教师，并注意在评价工作中

防止消极心理行为的产生.

练习和思考

1. 教育统计内容大致分哪两部分?
2. 作为一名教育工作者,为什么要学习教育评价与统计?
3. 作为一名教师应具备哪些教育评价与统计的知识和能力?

第一篇 教育统计

第一章 数据的初步整理

在教育领域内，无论是教育评价、教育实验还是调查研究，都需要采集大量的数字资料（数据）。只有通过对其数据进行整理、分析，才能使其特征和规律反映出来，才能根据数据所传递的信息进行科学推断，这个过程称为统计资料的整理。本章先介绍数据及其性质，然后着重介绍教育统计中对数据进行初步整理的常用方法。

§ 1 数据及其有关概念

一、数据和数据的特征

（一）数据

对客观事物用一定的方法进行测量，所得到的以数字形式表达的结果（比如，用分数表示的学生的考试成绩）或是通过调查得到的数字（比如，某区小学生人数）等，这些数字资料就叫做数据。

（二）数据的特征

通过测量所获得的数据，有以下基本特征：

1. 离散性。指观测到的大量数据以一定的分散形式出现。比如，对某个班或某个年级进行一次考试，学生的成绩分数不会全部相同，而是分散在一定的范围之内。

2. 变异性。指观测到的数据随时间和空间变化。在同样条件下，对同一事物多次观测得到的数据常常是有差异的。比如，一个学生的真实成绩是一定的，但多次进行同一水平的测验，各次

成绩一般不会完全一样，总是在真实成绩上下波动。

由于测量所得数据有变异性这一特点，在统计学上把它称为变量。像学生的身高、体重、成绩分数等数据都是变量。需要指出的是，在我们对研究对象进行测量之前，其取值是不能预料的，在统计学中把在测量之前不能预料数值的变量称为随机变量。

3. 规律性。指观测到的数据虽然有变异性，却不是杂乱无章的，而是在一定范围内呈现出某种规律。比如，对一个班或一个年级进行一次考试，所有学生的成绩分数看起来参差不齐，但将这些分数稍加整理就可以看出，得高分的人和得低分的人都比较少，分数居中的人数较多，这里呈现的就是“两头小、中间大”的规律。

正由于数据具有规律性这一特征，统计分析才有意义，才能使我们通过对数据的统计分析得到有价值的信息。

二、数据的种类

数据按不同的分类标准，可分为不同的种类，而不同种类的数据又有不同的统计分析方法。

（一）计数数据和测量数据

这是依据获得数据的来源和方法进行的分类。

1. 计数数据。又称枚举数据。是指以计算个数所获得的数据。如学生人数、教师人数、班级数、学校数等。

2. 测量数据。是指用一定的工具或按一定的标准，通过测量所获得的数据。这类数据又分三种。

（1）比率数据。又称比率变量。这种数据既有相等的单位，又有绝对零点。比如，以“米”为单位测量的学生身高，以“公斤”为单位测量的学生的体重等数据，都是比率数据。

比率数据之间可以用加、减、乘、除的运算来相互比较。例如，甲生身高 1.56 米，乙生身高 1.48 米，二者相比较，既可以说甲生比乙生高多少，也可以说甲生身高是乙生身高的几倍。

（2）等距数据。又称等距变量。这种数据只有相等的单位，但

无绝对零点。在教育测量中，以百分制计分的学生成绩分数，就属于等距数据，它只有大致相等的单位，而零分只是个相对值，对考试得零分的学生，不能认为他对所考科目一无所知。

等距数据之间只能用加、减运算相互比较，不能用乘、除运算。例如，在一次考试中甲生得 80 分，乙生得 60 分，我们只能说甲生比乙生高 20 分或乙生比甲生低 20 分；而不能说甲生得分是乙生得分的多少倍，因为这种说法是没有实际意义的。

(3) 等级数据。又称等级变量或顺序变量。这种数据既无相等单位，又无绝对零点。比如，以五分制计分的学生成绩分数、学生的操行评定等级、竞赛的名次等。

等级数据之间不能用加、减、乘、除的运算来进行比较。例如，一次考试甲生得 5 分，乙生得 4 分（均为五分制计分），两者相比，只能说甲生的成绩比乙生好，但不能说，甲生成绩比乙生高多少分或是乙生的多少倍，因为这些说法同样是没有实际意义的。

(二) 间断性数据和连续性数据

这是依据数据本身是否有连续性而进行的分类。

1. 间断性数据。是指有独立单位的数据，这种数据的两个单位之间不能再划分成细小的单位，数字形式一般是取整数。计数数据一般是间断性数据，如人数、班级数、学校数等。

2. 连续性数据。是指单位之间可以划分成无限多个细小单位的数据。比如，长度单位是“米”，它可以划分成分米、厘米、毫米、微米等。用一定工具测量而得到的数据，一般是连续性数据，像长度、质量、时间的数据等。

三、总体、个体、样本

(一) 总体

指具有某种特征的一类事物的全体，又称母体。总体是某一类研究对象的全体，总体的大小随研究的对象或问题而变。比如，我们的研究对象是在校的三年级学生，总体就是所有正在上三年

级的学生，而本校的三年级学生只是这个总体中的一部分；若我们的研究对象是本校的三年级学生，那么总体就是本校三年级的全体学生。

总体所包含的个体数目常用 N 表示。

(二) 个体

指构成总体的基本单位。比如，三年级学生是总体，则每个三年级学生就是一个个体。在教育统计中，个体不限于指人或物，也可以是每一个实验结果、每一个数据等。

总体的性质是由组成总体的每个个体决定的。

(三) 样本

指从总体中抽取的，用来观测的一部分个体的集合。比如，三年级学生是总体，本校的三年级学生就可看作是所有三年级学生的一个样本。

样本中所包含的个体数目的多少，称为样本容量，常用 n 表示。在教育统计中，一般将容量超过 30 的样本称为大样本，将容量等于或小于 30 的样本称为小样本。显然，样本越大，对总体的代表性就越强。当样本的大小不同时，统计的方法也不同，这将在后面讲到。

四、统计量和参量

(一) 统计量

又称样本统计量。指通过样本的数据计算出的可以描述样本统计特征的量数。比如，样本数据的算术平均值，用 \bar{X} 表示；样本数据的标准差，用 S 表示等。

(二) 参量

又称总体参量或总体参数。指描述总体（母体）情况的统计特征量数，也称母数。比如，总体算术平均值，用 μ 表示；总体标准差，用 σ 表示等。总体参量通常以希腊字母表示。

总体参量与样本统计量之间有着一定的关系。当总体参量未知时，样本统计量可以在某种程度上作为总体参量的估计值，这