

教育统计学

王汉澜 编著

教育科学出版社



教育统计学

王汉澜 编著

教育科学出版社

教育统计学

王汉澜 编著

责任编辑 金宏瑛

*

教育科学出版社出版

(北京北环西路10号)

新华书店北京发行所发行

中国科学院印刷厂印装

*

开本 850×1168 毫米 1/32 印张 11.75 字数 310,000

1986年11月第1版 1986年11月第1次印刷

印数 0,001—6,000 册

书号: 7232·267 定价: 2.60 元

8977-7

内 容 提 要

本书全面地系统地阐述了教育统计学的基本理论和基本方法，内容包括统计资料的搜集与整理、集中量数与差异量数、相关系数、回归分析、概率和常态分配、总体参数的估计、统计检验、方差分析、 χ^2 检验等，可供师范院校、教育学院作教材或教学参考书，也可供教育科学研究人员、教育行政管理人员学习参考。

编 者 的 话

本书是为教育系的教学需要而编写的，其目的是使学生比较系统而又全面地掌握教育统计学的基本理论和基本方法。本书也可供从事教育科学研究和教育行政管理工作的同志参考。

本书编写的基本原则是：以马克思列宁主义、毛泽东思想为指导，力求作到科学性和思想性统一；汲取国外教育统计学的先进成果，力求理论联系实际，适合我国教育情况；从教育统计工作的实际需要出发选编内容和安排章节顺序，删除不实用的统计方法。本书先介绍统计资料的搜集整理、统计图表的制作，后讲各种统计的方法，把描述统计与推断统计有机地结合起来，由浅入深，前后连贯，力求便于教学；在概念的表述、理论的阐述和方法的论证上，力求明晰准确、步骤清晰，以便读者阅读。

本书在编写的过程中，参考了国内外有关的书籍和资料，从中吸取了一些材料，特此致谢！

本书曾于 1982 年印出了初稿，现根据两年的教学实践经验，作了一些修改，在修改过程中，扈涛同志作了一些协助工作。诚恳地希望读者对本书批评指正！

王汉澜

一九八五年四月

目 录

第一章 绪论	1
第一节 教育统计学的研究对象.....	1
第二节 教育统计学的基本术语.....	4
第三节 教育统计学的研究方法.....	7
第四节 学习教育统计学的重要意义.....	9
第二章 统计资料的搜集	12
第一节 统计资料的来源和种类.....	12
第二节 统计资料搜集的途径和方法.....	13
第三节 对搜集资料的要求.....	21
第三章 统计资料的整理	24
第一节 教育统计资料整理的意义及其分类.....	24
第二节 数据资料的整理.....	26
第四章 教育统计表和统计图	32
第一节 教育统计表.....	32
第二节 教育统计图.....	38
第五章 集中量数	55
第一节 算术平均数.....	55
第二节 中数和众数.....	65
第三节 几何平均数和调和平均数.....	71
第四节 各种集中量数的比较.....	78
第六章 差异量数	81
第一节 差异量数的意义和种类.....	81
第二节 两极差.....	82
第三节 四分差.....	84
第四节 平均差.....	87

第五节	标准差	90
第六节	相对差异量数	101
第七节	相对位置量数	103
第八节	各种差异量数的比较	110
第七章	相关系数	113
第一节	相关的意义和种类	113
第二节	积差相关	119
第三节	等级相关	130
第四节	点双列相关	133
第八章	回归分析	138
第一节	回归分析的意义	138
第二节	回归系数的计算	143
第三节	回归值的计算	147
第四节	回归值的标准误差	151
第九章	概率和常态分配	158
第一节	概率	158
第二节	常态分配	166
第三节	偏态分配	179
第四节	t 分配	182
第十章	总体参数的估计	193
第一节	总体参数估计的意义	193
第二节	总体参数估计的理论基础	196
第三节	总体平均数的估计方法	202
第四节	总体标准差的估计方法	206
第五节	总体百分比的估计方法	208
第六节	总体相关系数的估计方法	209
第十一章	统计检验	215
第一节	统计检验的意义和一般步骤	215
第二节	平均数的差异显著性检验	221
第三节	标准差的差异显著性检验	235

第四节	相关系数的差异显著性检验	239
第十二章	方差分析	245
第一节	方差分析的意义	245
第二节	单因素方差分析与 F 检验	250
第三节	F 检验后的 t 检验和方差齐性检验	262
第四节	双因素的方差分析与 F 检验	265
第十三章	χ^2 检验	283
第一节	χ^2 检验的意义和方法	283
第二节	χ^2 检验的应用	290
附录表:	某市 04 中学 80 年级学生原始数据	302
	书中统计符号集注	364

第一章 绪 论

第一节 教育统计学的研究对象

一、统计与统计学

要了解什么是教育统计学,须先了解什么是统计和统计学。

统计是总括起来计算的意思,它是指的对某一现象的有关资料的搜集、整理、计算和分析的工作过程。例如,将我国建国三十多年来有关教育事业发展的资料,加以搜集、整理、计算,得出我国现有普通中学 118377 所,比解放前最高年增长 26.7 倍。现有普通中学学生 5508.1 万人,比解放前最高年增长 35.8 倍。这一工作过程,就叫做统计或统计工作。

统计工作,在人类历史上早已出现,古代的国家,常用它来了解、研究人口、土地、物产、财富、兵力等问题。如《禹贡九州篇》中就记载了当时我国各地的土质、河山、物产、田赋等,这是运用统计方法编写的一本巨著,在古希腊、罗马和埃及各国,也是用统计的方法来考查国家的人口、财富等政治情况的。

随着资本主义工商业的发展和数学理论的发展,统计的使用范围逐渐扩大,统计的方法日益增多。十七世纪中叶,法国人帕斯卡(Pascal 1623~1662 年)和佛尔麻梯(Fermat 1601—1665 年)始创概率论,为统计学的发展提供了一定的基础。一七三三年法国人毛弗尔(De Moivre 1667—1752 年)最先发现了正态曲线,只不过当时未得到应用。一八二六年德国数学家高斯(Gauss 1777~1855 年)在阐述观察值的误差理论时,又导出正态曲线。正态曲线的发表,为大量的数据统计工作,创造了有利的条件。而后,英国生物学家高尔顿(F. Galton 1822—1911 年)创立了回归线的原理,皮尔逊(Karl Pearson 1857—1936 年)发表了直线相关系数 r 的用法。英国人高赛特(W. S. Gosset 1876—1937 年)研究出了小

样本的统计方法,美国数学家费舍(R. A. Fisher 1890~1962年)发表了Z函数与F分配等,并于1925年左右出版了他的《研究工作者用数理统计方法》一书;至此统计学便成为一门独立的科学。

统计学是研究数字资料的搜集、整理、计算和分析的方法及其原理的一门科学,它是通过对大量的自然现象和社会现象的数量研究,来反映客观事物、现象的特征及其规律性的一门科学。

一切事物无不具有质量与数量两个方面,这两个方面是辩证统一的。我们在研究事物、现象时,首先要了解事物、现象的属性,查明它与其他事物、现象的区别,确定事物、现象的质量。但同时,也要注意事物、现象的数量。毛泽东同志曾经指出:“对情况和问题一定要注意到它们的数量方面,要有基本的数量的分析。任何质量都表现为一定的数量,没有数量也就没有质量。我们有许多同志至今不懂得注意事物的数量方面,不懂得注意基本的统计、主要的百分比,不懂得注意决定事物质量的数量界限,一切都是胸中无‘数’,结果就不能不犯错误。”^①因此,我们在研究事物、现象时,必须注意数量分析,而数量分析的科学方法,就是统计学所要研究的问题。

无论是自然现象的规律或是社会现象的规律,都是通过大量的事实数据表现出来的。如有计划、按比例发展的社会主义经济规律,是通过大量的积累、消费、农业、轻工业和重工业等方面的数字,才能反映出它们之间的比例关系及其发展的规律性,个别方面的少量数字是不能准确地反映出比例关系及其发展规律的。再如,社会主义国家国民收入的分配特征,是通过大量的职工的工资和农民的收入,从而计算出平均数和标准差,才能反映出来。个别职工的工资或农民收入,可能受某种偶然因素的影响,不足以说明社会主义国家国民收入的分配特征。所以,当我们研究社会主义国家国民收入的分配问题时,不能仅仅依据少数的或个人的收入,而应根据大量的国民收入的数字来进行研究。否则,所得的结果是不准确的、不可靠的。只有在大量的事实或数据的情况下,个

^① 《毛泽东选集》第四卷,人民出版社1960年版,第1443页。

别的事实或数据的作用,才可以相互抵消或削弱,不影响或很少影响总的结果。所以,统计学所要研究的是大量的数据,要从对大量的数据的分析研究中,揭示出事物、现象的特征和规律性。

就统计学所研究的方法来说,统计学可分为描述统计和推断统计两大类:

描述统计:这类统计的方法是对研究的问题进行全面的调查或观测,将所得的全部数据,加以归类、分组,简缩成易于处理的形式,用表列法、图示法反映这些数据的分布特征,计算出具有代表性的统计数字,如集中量数、差异量数和相关系数等。从而反映出这些数据在某一方面的特征与相互关系。如计算全班学生成绩的平均分数,语文成绩与数学成绩的相关系数等。这类统计方法在社会科学和实际工作中应用最广。

推断统计:推断统计又叫抽样统计,是在描述统计的基础上发展起来的,是用抽样的方法,从总体中抽出一部分个体,作为样本,从对样本的研究中得出的统计数字(统计量),来推断总体的有关特征。或者从一个已知量数去估计、推断一个相应的未知量数,也都是推断统计的内容。这类统计方法是近三、四十年发展起来的。

就统计学所研究的对象的性质来说,统计学可以分为数理统计学和专业统计学。

数理统计学(理论统计学、统计学原理)是研究统计的一般原理和方法。它的主要任务是阐明统计的原理、原则、公式。它不专门阐述某一领域内的统计问题。

专业统计学是把统计学的一般原理和方法应用于某一专业,又称应用统计学。它的主要任务不是阐明统计的原理和公式,而是研究如何把统计的方法运用于某一专业。如经济统计学、生物统计学、卫生统计学、教育统计学等。

二、教育统计学的研究对象

教育统计学是运用统计学的原理和方法来研究教育问题的一门应用科学,它是研究教育现象中数量关系的一门科学,通过对教

育现象的数量分析,来帮助我们认识教育的现象和掌握教育规律。也可以说,教育统计学是研究如何搜集、整理、计算和分析教育上的数字资料,从而表明教育上的某些特征及其规律性的一门科学。

统计学是以大量的事实或数据为研究对象的,进一步来说,是以群体现象为研究的对象,教育统计学也是以教育上的群体现象为研究对象。所谓群体,是对个体而言;所谓群体现象是对个别现象和典型现象而言。例如,一个人是一个个体,许多人就是人的群体。再如,许多人组成一个社团,许多社团又组成一个社团的群体,每个社团对社团群体来说又是一个个体。统计学的研究对象是群体现象。也就是说,教育统计学要从教育上的大量的数字资料中去研究问题。不对个别的数字资料进行研究。例如,某校某生某门学科某一次考试成绩为95分,这不能构成统计学研究的对象,因为这一成绩,既不能代表该校该班学生的一般水平,也不足以说明该生该门学科成绩的发展情况。如果研究的是很多学生的一次考试成绩或一个学生的多次考试成绩,那就成了教育统计学的研究对象。

教育统计学的一般结构,可分为三部分:

第一部分是统计资料的搜集。主要是论述如何全面而又准确地搜集客观的数字资料。

第二部分是统计资料的整理。主要是阐明如何整理资料、计算综合指标、绘制统计图表。

第三部分是统计资料的计算和分析。主要是研究如何计算出统计资料的各种具有代表性的统计数字。包括有:集中量数、差异量数、相关系数、回归分析、总体参数的估计和差异的显著性检验等。这是教育统计学的主要部分。

第二节 教育统计学的基本术语

在统计学和教育统计学中,经常使用一些专门的术语,如果不

了解这些术语,就很难理解它们的内容。现将教育统计学中使用的基本术语及其含义,介绍于下:

一、数据和数列

在统计工作中,对由调查和实验得来的数字资料,叫做数据。

许多数据经过分类,将同性质的数据,依照某种法则排列的一些数据叫做数列。例如:自然数列: $1, 2, 3, \dots, n, \dots$; 偶数列: $2, 4, 6, \dots, 2n, \dots$ 。数列有许多种,在教育统计中常见的有如下几种:

1. 时间数列: 按照统计事项的数据的先后时间顺序而排列的数列,叫做时间数列。如将河南省每年的小学生人数,按年代的顺序排列,即称之为时间数列。

2. 空间数列: 按照统计事项的数据的空间顺序排列的数列,叫做空间数列。如将河南省各县小学生的人数,按地区、县份加以排列,即称之为空间数列。

3. 品质数列: 按照统计事项的性质或等级而排列的数列,叫做品质数列。如学生的操行甲、乙、丙、丁; 人的皮肤颜色: 白、黄、棕、黑等。

4. 数量数列: 按照统计事项数量的大小顺序而排列的数列,叫做数量数列。数量数列中有连续数列和间断数列两种,凡所列的数值单位能够作无穷的精微划分,在任两个数之间可以想象出有无限数值介于中间,这样的数列是连续数列,反之,则为间断数列。如学生的成绩由 0 至 100 分之间,可以有許多连续的分數,这种数列叫连续数列。人数 1, 2, 3, 4 中间,不能再有数值介于中间,叫做间断数列。

二、总体与个体

总体又称母体、母全体,它是对个体或子体而言。在统计学中把所研究的数据的全体,叫做总体。其中一个单位或一部分称

为个体或子体。如我们要研究全校二年级学生的语文成绩，全校二年级学生的语文成绩称为总体，其中一班或一个学生的语文成绩，称为子体。

三、抽样与样本

在统计工作中，往往所研究的总体所包含的数目很多，不宜全部拿来进行研究，例如，我们要研究全国高中毕业生的数学程度，把这一总体的各个学生的成绩都拿来研究，是非常麻烦与困难的，我们就从中抽取一部分学生的数学成绩作为样本来研究，这种从总体中取出一部分个体的办法，称为抽样或取样。所抽得的部分就称之为样本。样本里所含有的个体的数目，叫做样本的含量或样本的容量。

四、变量与常数

我们统计学中所研究的数量，是以一个个的数值形式出现的，在某一过程中(或一定条件下)可以出现不同数值的统计事项，叫做变量，如学生的年龄、身高、体重、学习分数等，称之为变量。数值保持不变的量，叫做常量或常数。如 $\sqrt{2}$ 和圆周率 π ，则是常数。

五、自变量与因变量

变量依其相互关系，可以分为自变量与因变量。在一问题中有两个量 x 、 y ，如果对量 x 所取的每一个值，都对应着 y 的唯一确定的值，就称量 y 是量 x 的函数。量 x 叫做自变量，量 y 是量 x 的函数，或说量 y 是因变量。

六、品质变量与数量变量

变量依其性质不同，可分为品质变量和数量变量。在统计的资料中，有些资料是属于说明某一事物与其它事物在属性上、类别上或等级上不同的，不是表示事物之间差异的大小，这样的变量叫做品质变量。如学生的性别分男、女，人的皮色有黄、白、黑，学生

的品行有好、坏,或甲、乙、丙等。有些统计资料的性质相同,而表示大小数量不同的,这样的变量,叫做数量变量。

七、统计量与参数

在统计工作中,对一系列的数据进行计算,计算出的平均数、标准差及其它测度,叫统计量。如对 50 个中学生的数学进行测验,对测验得来的 50 个数据进行计算,得出一个平均分数,这个平均数就是一个统计量。一般在统计学中,从原始量数直接计算出样本的平均数、标准差及其它测度,叫做统计量。由样本的统计量而估计、推断出来的相应的总体的统计量,称为参数或参量。

在统计学中还有许多术语,如随机取样、自由度、显著性水平等,我们将在下面有关章节中介绍。

第三节 教育统计学的研究方法

教育统计学在研究问题时,经常采用以下几种方法:

一、大量观察法

大量观察法是搜集统计资料的一种基本方法,它是对研究对象的全部或足够数量的事实进行调查研究,从而显示整个现象的一般特征或发展趋势的一种方法。

教育统计学所以要采用大量观察法,一是因为客观事物的规律性是通过大量的现象才表现出来,二是由于教育统计学所要研究的问题是群体现象,只有通过大量的观察,才能认识所研究对象的总的情况和趋势。

二、类型分组法

类型分组法是整理统计资料的一种基本方法,它是把大量的统计资料,按照事物现象的类别、性质、数量、程度等,划分出不同的类别和组别,以便进一步地进行分析研究。

教育现象是错综复杂的,如果没有把不同性质、类别的东西区分开来,就会混淆是非,难以弄清它们的特征及其规律性;在大量的性质相同的现象中往往在量上或程度上也有所不同,不加分组,也不易进行研究,所以在教育统计学中要采用类型分组法。列宁非常重视类型分组,他说:“如果没有全面的、编制得合理的分组和复合表,极丰富的按户调查资料简直毫无用处。”^①又说:“统计工作不是把数字随便填到几个格格里去,而应当是用数字来说明所研究的现象在实际生活中已经充分呈现出来或正在呈现出来的各种社会类型。”^②在教育统计学中经常采用类型分组法来区分教育现象的类型,反映教育现象的内部结构,揭示教育现象之间的关系。如将各级各类学校的学生人数按学校类型来加以划分,进行统计;或再将性别不同的男女生分别进行统计等。

三、综合指标法

综合指标法是对已整理过的统计资料,用各种综合指标来概括地表明所研究对象的一般特征的方法,它的基本形式是绝对数、相对数和平均数,如某校共有30个班,学生1200人,这是综合指标的绝对数;其中男生占60%,女生占40%,这是综合指标的相对数;平均每班有学生40人,这是综合指标的平均数。

四、统计图表法

统计图表法是把已经整理好的数据用图或表的形式把它显示出来,起到直观、醒目的作用,也便于进行比较、对照和分析研究。

五、统计检验法

统计检验法是对由教育调查得来的数据或教育实验的结果,予以科学的处理、分析和解释。这种方法包括有Z检验、t检验、

① 《列宁全集》第20卷,第72页。

② 《列宁全集》第18卷,第254页。

F 检验、 χ^2 (卡方) 检验。

以上各种方法都是教育统计工作中不可缺少的方法，它们是互相联系的，如果在搜集资料时不科学，计算的再正确，也会得出错误的结论。如果搜集的资料科学，而计算的方法，检验的方法不对，也不能反映事物的特征。所以，我们在研究问题时，应把它们结合起来使用，以便全面而又准确的反映研究对象的特征，揭示其发展的趋势及其规律性。

第四节 学习教育统计学的重要意义

学习教育统计学，对于教育科学工作者和教育实际工作者，都具有十分重要的意义。

首先，教育统计学能帮助我们认识教育现象和了解教育的规律。

统计是认识社会、管理国家和处理工作的一种最科学的工具。社会上人口的多少，产品的类别和数量，商品流通的情况，教育事业的规模和发展速度等等，只有通过统计，才能得到了解和说明。国民经济的发展规划，产品分配的核算，教育事业的管理，社会工作的处理，无不需统计，列宁曾经指出：“有许多问题，而且是涉及到现代国家的经济制度和这种制度的发展的最根本问题，过去是根据一般的估计和大致的材料加以解决的，现在如果不根据按某一个一定的纲要收集并经统计专家综合的关于某一国家全国情况的浩繁材料，就无法加以比较认真的研究。”^① 斯大林也曾说过：“任何建设工作，任何国家工作，任何计划工作，如果没有正确的计算是不可想象的。而没有统计，计算是不可想象的。没有统计，计算就一步也不能前进。”^②

其次，教育统计是制订教育事业发展规划和检查教育工作的重要手段之一，是教育行政部门的一项重要工作。

① 《列宁全集》第4版，第16卷，第420页。

② 《斯大林全集》第6卷，第189页。