



普通高等教育
“十一五”国家级规划教材



教育类专业基础课系列教材

第五版

教育统计学

王孝玲◎著



华东师范大学出版社



普通高等教育
“十一五”国家级规划教材



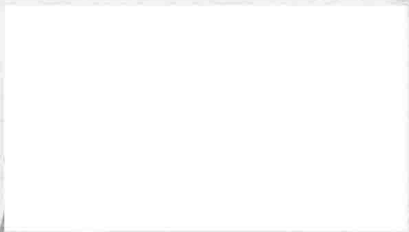
教育类专业基础课系列教材

第五版

教育统计学

王孝玲◎著

G64-05 /
1114.1



华东师范大学出版社

图书在版编目(CIP)数据

教育统计学/王孝玲著. —5版. —上海:华东师范大学出版社,2014.9

教育类专业基础课系列教材

ISBN 978-7-5675-2554-2

I. ①教… II. ①王… III. ①教育统计—统计学—高等学校—教材 IV. ①G40-051

中国版本图书馆CIP数据核字(2014)第217135号

教育类专业基础课系列教材

教育统计学(第五版)

著 者 王孝玲
责任编辑 吴海红
审读编辑 宋亚洲
装帧设计 卢晓红

出版发行 华东师范大学出版社
社 址 上海市中山北路3663号 邮编 200062
网 址 www.ecnupress.com.cn
电 话 021-60821666 行政传真 021-62572105
客服电话 021-62865537 门市(邮购)电话 021-62869887
地 址 上海市中山北路3663号华东师范大学校内先锋路口
网 店 <http://hdsdcbs.tmall.com>

印 刷 者 昆山市亭林彩印厂有限公司
开 本 787×1092 16开
印 张 23.5
字 数 584千字
版 次 2015年2月第5版
印 次 2015年2月第1次
书 号 ISBN 978-7-5675-2554-2/G·7624
定 价 49.00元

出版人 王 焰

(如发现本版图书有印订质量问题,请寄回本社客服中心调换或电话021-62865537联系)

第五版 前言

该书自1986年出版至今,28年来每年从未间断过印刷。据《中国教育报》2004年1月1日第5版报道,由全国教师用书发行协会的统计,在全国高校教育统计学教材发行量排行榜上名列第一。该书修订二版获得了2003年度上海市优秀教材奖。2006年,该书被列为普通高等教育“十一五”国家级规划教材。2009年该书第四版荣获中国大学出版社图书奖首届优秀教材奖二等奖。

为了适应日益发展的教育科研数量化分析的需要,该书必须更新,即增加新的而去掉旧的统计方法和计算手段。本着这一原则此次作如下修订:

增加了常用的多元统计方法“主成分分析”、“聚类分析”两章,使多元统计的内容更加充实、完整、齐全;重新编排和调整了多元统计方法的章节次序,使其组织结构更加合理、顺畅、符合逻辑。

删去了较为陈旧的统计方法及已经过时的统计工具的使用;减化了不常用的冗长的统计附表;合并了重复的计算过程;紧缩了公式及计算式的表达形式;去掉了公式的推导及绝大部分证明(因为这是数理统计学的任务);精减了对教育统计学来说不甚重要的理论与概念。

在本次修订中,第十五章第三节和第十六章第四节,由中国浦东干部学院王君副教授撰写。衷心感谢许雷同志在本书的初版及几次修订中给予的帮助!

敬请同行、读者对此次修订版中的错误及不妥之处给予批评指正。

王孝玲

2014年1月26日

第四版 前言

该书于1986年出版以来,据《中国教育报》2004年1月1日第5版报道,由全国教师用书发行协会的统计,在全国高校教育统计学教材发行量排行榜上名列第一。该书修订二版获得了2003年度上海市优秀教材奖。2006年,该书被列为普通高等教育“十一五”国家级规划教材。

本次修订主要在以下几个方面:一是增加了多元统计方法中常用的“因素分析”一章,以适应日益发展的教育科研数量化分析的需要;二是全面系统地修改了符号系统,使之更清晰地表达统计量;三是更换了某些例子,使其更接近教育科研实际,应用时更具有可模仿性;四是引导读者综合地运用统计方法分析教育问题。

在本次修订中,第十五章第四节,由中国浦东干部学院王君副教授撰写。衷心感谢许敏、许雷两位同志在本书的初版及几次修订中给予的帮助!

限于作者的水平,此次修订版可能还会有错误或不妥之处,敬请同行、读者批评指正。

王孝玲

2007年8月23日

第四章 量数	103
第一节 全距、四分位数、百分位数	103
第二节 平均数	105
第三节 方差和标准差	106
第四节 相对量数	109
第五节 标志量及标志差	110
第五章 概率及概率分布	116
第一节 概率的一般概念	116
第二节 二项分布	118
第三节 正态分布	124
第六章 抽样分布及总体平均数的推断	134
第一节 抽样分布	134
第二节 总体平均数的估计	137
第三节 假设检验的基本原理	144

目 录

第一章 绪论	(1)
第一节 什么是教育统计学	(1)
第二节 教育统计学中的几个基本概念	(2)
第二章 数据的初步整理	(4)
第一节 数据的来源、种类及其统计分类	(4)
第二节 统计表	(6)
第三节 统计图	(11)
第三章 集中量	(19)
第一节 算术平均数	(19)
第二节 中位数	(21)
第三节 众数	(24)
第四节 加权平均数、几何平均数、调和平均数	(27)
第四章 差异量	(32)
第一节 全距、四分位距、百分位距	(32)
第二节 平均差	(35)
第三节 方差和标准差	(36)
第四节 相对差异量	(39)
第五节 偏态量及峰态量	(40)
第五章 概率及概率分布	(46)
第一节 概率的一般概念	(46)
第二节 二项分布	(48)
第三节 正态分布	(54)
第六章 抽样分布及总体平均数的推断	(64)
第一节 抽样分布	(64)
第二节 总体平均数的估计	(67)
第三节 假设检验的基本原理	(74)

第四节 总体平均数的显著性检验 (77)

第七章 平均数差异的显著性检验 (83)

第一节 平均数差异显著性检验的基本原理 (83)

第二节 相关样本平均数差异的显著性检验 (84)

第三节 独立样本平均数差异的显著性检验 (90)

第四节 方差不齐性独立样本平均数差异的显著性
检验 (94)

第五节 方差齐性检验 (97)

第八章 方差分析 (103)

第一节 方差分析的基本原理 (103)

第二节 完全随机设计的方差分析 (107)

第三节 随机区组设计的方差分析 (112)

第四节 各个平均数差异的显著性检验 (115)

第五节 多组方差的齐性检验 (119)

第六节 多因素方差分析简介 (122)

第九章 总体比率的推断 (129)

第一节 比率的抽样分布 (129)

第二节 总体比率的区间估计 (130)

第三节 总体比率的假设检验 (132)

第四节 总体比率差异的显著性检验 (134)

第十章 χ^2 检验 (138)

第一节 χ^2 及其分布 (138)

第二节 单向表的 χ^2 检验 (140)

第三节 双向表的 χ^2 检验 (144)

第四节 四格表的 χ^2 检验 (147)

第十一章 相关分析 (153)

第一节 相关的意义 (153)

第二节 积差相关 (154)

第三节 等级相关 (164)

第四节 质与量的相关 (169)

第五节 品质相关 (176)

第十二章 回归分析 (184)

第一节 一元线性回归 (184)

第二节	一元线性回归方程的检验	(188)
第三节	一元线性回归方程的应用	(193)
第四节	多元线性回归简介	(196)
第十三章	非参数检验	(204)
第一节	符号检验	(204)
第二节	符号秩次检验	(208)
第三节	秩和检验	(211)
第四节	中位数检验	(214)
第五节	单向秩次方差分析	(216)
第六节	双向秩次方差分析	(218)
第十四章	抽样设计	(223)
第一节	抽样方法	(223)
第二节	总体平均数统计推断时样本容量的确定	(225)
第三节	总体比率统计推断及相关系数显著性 检验时样本容量的确定	(231)
第十五章	主成分分析	(235)
第一节	主成分分析的基本原理	(235)
第二节	样本主成分	(236)
第三节	应用举例	(240)
第十六章	因素分析	(248)
第一节	因素分析的基本原理	(248)
第二节	因素负荷矩阵的求法	(251)
第三节	旋转变换及因素计分	(254)
第四节	应用举例	(258)
第十七章	聚类分析	(262)
第一节	距离与相似系数	(262)
第二节	系统聚类法	(268)
第三节	应用举例	(282)
	练习题答案	(286)
	统计用表	(300)
附表 1	正态分布表	(300)
附表 2	t 值表	(304)
附表 3	F 值表	(306)

附表 4	q 值表	(314)
附表 5	F_{\max} 界值表(哈特莱方差齐性检验)	(316)
附表 6	百分率的可信限	(317)
附表 7	χ^2 值表	(331)
附表 8	r 值的 Z_r 转换表	(333)
附表 9	积差相关系数界值表	(335)
附表 10	等级相关系数界值表	(339)
附表 11	复相关系数界值表	(341)
附表 12	符号检验表	(343)
附表 13	符号秩次检验表	(344)
附表 14	秩和检验表	(345)
附表 15	H 检验表	(346)
附表 16	双向秩次方差分析 χ^2 值表	(348)
附表 17	随机数字表	(351)
附表 18	由样本平均数估计总体平均数所需样本容量 n	(352)
附表 19	样本平均数与总体平均数差异显著性检验所需样本容量 n	(354)
附表 20	两个样本平均数差异显著性检验所需样本容量 $n(=n_1=n_2)$	(356)
附表 21	由样本比率估计总体比率所需样本容量 n	(358)
附表 22	$\sqrt{\text{比率}}$ 的反正弦转换表	(360)
附表 23	两个样本比率差异显著性检验所需样本容量 $n(=n_1=n_2)$	(361)
附表 24	相关系数显著性检验所需样本容量 n	(363)

第一章 绪论

第一节 什么是教育统计学

一、教育统计学的任务

教育统计学是运用数理统计的原理和方法研究教育问题的一门应用科学。它的主要任务是研究如何搜集、整理、分析由教育调查和教育实验等途径所获得的数字资料,并以此为依据,进行科学推断,从而揭示蕴含在教育现象中的客观规律。但是从研究内容来说,教育研究课题的提出,内容的界定,对象范围的确定,假设的建立,结论的得出以及分析,却不是教育统计学的研究任务,因为这些问题还要依靠与研究内容有关的教育专业知识来解决。而教育统计学只能提供各种统计方法的应用条件和统计计算结果的解释。另外,统计原理和方法的数学证明及公式推导,也不是它的主要任务,而是数理统计学的任务。

二、教育统计学的内容

教育统计学的研究内容,从具体应用角度来分,可以分成描述统计、推断统计和实验设计三部分。由于实验设计比较复杂,可以独立成为一门学科,故一般教育统计学以阐述描述统计和推断统计两部分内容为主。

1. 描述统计

对已获得的数据进行整理、概括,显现其分布特征的统计方法,称为描述统计。通过教育调查和教育实验等获得了大量的数据,用归组、编表、绘图等统计方法对之进行归纳、整理,以直观形象的形式反映其分布特征;通过计算各种特征量,来反映它们分布上的数字特征。例如,计算集中量(如算术平均数、中位数、众数、加权算术平均数、几何平均数、调和平均数等)来反映它们的集中趋势;计算差异量(如全距、四分位距、百分位距、平均差、标准差、差异系数等)来反映它们的离散程度;计算偏态量及峰态量来反映它们的分布形态;计算相关量(如积差相关系数、等级相关系数、点二列相关系数、二列相关系数、 Φ 相关系数、四分相关系数、C相关系数、肯德尔和谐系数、多系列相关系数等)来反映两个或多个变量之间变化的一致性程度。这些均属于描述统计范围。其目的在于将大量零散的、杂乱无序的数字资料进行整理、归纳、简缩、概括,使事物的全貌及其分布特征清晰、明确地显现出来。

2. 推断统计

根据样本所提供的信息,运用概率的理论进行分析、论证,在一定可靠程度上对总体分布特征进行估计、推测,这种统计方法称为推断统计。推断统计的内容包括总体参数估计和假设检验两部分。例如,对总体参数值,即总体数字特征值(如总体平均数、总体标准差、总

体相关系数等)的估计;对总体参数或总体参数之差(如总体平均数之差、总体方差之差、总体相关系数之差等)的假设检验;对总体分布是否服从某种分布的假设检验,等等,都属于推断统计的范围。其目的在于根据已知的情况,在一定概率意义上估计、推断未知的情况。

三、教育统计学是教育科研定量分析的重要工具

凡是客观存在的事物,都有数量的表现。凡有数量表现的事物,都可以进行测量。教育现象是一种客观存在的事物,它也有数量的表现。虽然教育现象具有多变性,而且引起它发生变化的因素甚多,难以准确地测量,但是它毕竟还是可以测量的。因此,在进行教育科学研究时,在一定条件下,是可以对教育现象进行定量分析的。教育统计学就是对教育问题进行定量分析的重要科学工具。无论是教育调查,还是教育实验,都需要用统计方法进行处理和分析。特别是教育实验,步步都离不开统计的方法和原理。例如,实验的设计,被试的选择,样本容量的确定,实验因子的安排,无关因素的控制,数据的整理、分析,统计的推断,结果的表述和解释等,各个环节都必须对统计方法有较深刻的理解和熟练的掌握,才能顺利进行。

第二节 教育统计学中的几个基本概念

一、随机变量

为了解释随机变量的概念,先介绍随机现象和随机事件。

具有以下三个特性的现象,称为随机现象。第一,一次试验有多种可能结果,其所有可能结果是已知的;第二,试验之前不能预料哪一种结果会出现;第三,在相同的条件下可以重复试验。例如,抛一枚硬币,有两种可能结果:不是正面朝上,就是反面朝上;究竟哪面朝上,事先不能预料;相同的条件下可以重复抛多次。这种现象是随机现象。随机现象的每一种结果叫做一个随机事件。这些随机事件在一次试验中,可能出现,也可能不出现,而在大量重复试验中,它们的发生却具有一定的规律性。例如,硬币的正面朝上称为随机事件 A ,反面朝上称为随机事件 B 。在抛一次硬币时,事件 A 可能发生,也可能不发生,但如果重复抛许多次,事件 A 的发生就会具有某种规律性,即它出现的频率接近 $1/2$ 。我们把能表示随机现象各种结果的变量称为随机变量。统计处理的变量都是些随机变量。例如,学生的身高、体重、性别、智商、某科考试成绩,教师的人数、年龄、教龄、工资,等等。本书用大写英文字母 X 、 Y 等,分别表示不同的随机变量。例如,用 X 表示某班语文测验分数这一随机变量;用 Y 表示数学分数这一随机变量。而该班每个学生的语文分数可以分别用 X_1, X_2, \dots, X_n 或 X_i 来表示;数学分数可以分别用 Y_1, Y_2, \dots, Y_n 或 Y_i 来表示。每个随机事件往往表现为一种数值。对于不是以数值表示的随机事件,可以将之数量化。例如,可将高考录取和未录取分别用1和0表示,将品德评定的优、良、中、差等级分别用4、3、2、1表示。

二、总体和样本

总体是我们所研究的具有某种共同特性的个体的总和。总体中的每个单位称为个体。我们所要研究的往往是这些个体某一方面的特性,如学生的身高、体重,某种知识、能力,作

业所用的时间,等等。但是总体中的个体,不一定是人的某种属性。假如,我们要考查小学生的识字量,这时,从测验材料角度来说,所有的常用字就是一个总体,而其中的每个字就是个体。样本是从总体中抽取的作为观察对象的一部分个体。当对总体某种特性进行研究时,限于人力、物力、经费和时间,不可能将总体中的每一个个体一一进行观测,往往需要从中抽取一部分个体,作为样本进行观察、分析,然后根据样本所获得的信息,在一定可靠度上推断总体。

当总体所包含的个体数目有限时,这一总体称为有限总体。而总体所包含的个体数目无限时,则称为无限总体。例如,我们研究某区高三英语毕业会考成绩,这是个有限总体。当我们研究8岁女童的身高,以古代人、现代人、中国人、外国人作为测查对象时,则这里的8岁女童为无限总体。从理论上说,如果我们对某个地区8岁女童的身高(有限总体)进行无限次的测量,则测量的一切可能结果,也可形成一个观察值上的无限总体。有限总体内所包含的个体数目,一般用 N 来表示。在实际研究工作中,总体应当选择有限的还是无限的,以及对于有限总体来说,总体内应当包含多少个体,这都应依研究的问题所欲推断的范围而定。例如,从某区随机抽取6所学校,对其学生家长的职业进行调查。此时,这6所学校的学生家长,可以作为该区的样本,同时也可以作为这6所学校的总体。总体和样本是相对的。当然,总体内的个体要有相当大的数量。以一个班的几十个人作为一个总体,未免太小。

样本中包含的个体数目称为样本的容量,一般用 n 表示。样本中个体数目大于30一般称为大样本,等于或小于30称为小样本。在有些情况下对数据进行统计处理时,大样本和小样本所用的统计方法并不相同。关于从总体中用什么方式抽取样本,样本容量如何确定,将在后面有关章节中加以叙述。

三、统计量和参数

样本上的数字特征是统计量。也就是说,根据实得的数据所计算出的能够描述这组数据各种特征的数量是统计量。例如,描述一组数据集中趋势的一种统计指标称为平均数(用 \bar{X} 或 \bar{Y} 表示);描述一组数据分散程度的一种统计指标称为标准差(用 σ_x 表示);描述某一事物两种特征之间关系的统计指标称为相关系数(用 r 表示),等等。这些都是统计量。

总体上的各种数字特征是参数。也即反映总体上各种特征的数量是参数。例如,反映总体集中趋势的一种统计指标称为总体平均数(用 μ 表示);反映总体内个体间分散程度的一种统计指标称为总体标准差(用 σ 表示);反映某一事物的两种特征之间在总体内变化关系的一种统计指标称为总体相关系数(用 ρ 表示),等等。这些都是总体参数。

在进行统计推断时,就是根据样本统计量来推断总体相应的参数。如根据样本的平均数推断总体的平均数;根据样本的标准差推断总体的标准差;根据样本的相关系数推断总体的相关系数;或者根据样本某种统计量指标的差数,推断总体相应指标差数的参数。

练习 题

1. 什么是教育统计学? 它的研究任务及主要内容是什么?
2. 学习教育统计学有什么重要意义?
3. 什么叫随机变量、总体与样本、统计量与参数?

第二章 数据的初步整理

第一节 数据的来源、种类及其统计分类

统计所处理的是数字资料。搜集数字资料是统计工作的第一步,而且是统计整理和分析的基础。如果搜集的资料不准确、不完整、不系统,那么不仅资料本身无应用价值,而且以此为依据所进行的统计整理和分析,会导致错误的结论。因此,搜集资料是统计工作的重要一环。

一、教育统计资料的来源

教育统计资料的来源有两个方面。

1. 经常性资料

主要指文字记载的资料,包括日常工作记录和统计报表等。例如,学校中关于老师的年龄、教龄、职称、工资、教育和教学工作情况的记录,关于学生的学期和学年各科成绩、品德评定、健康状况、奖惩情况、家长职业的记录,学校经费使用的记录,学制、课程、教材、教法改革情况,以及教育科学研究成果的记录,等等。统计报表主要是指教育部门逐级向上呈报的各种表格,如教职工登记表,在校各年级学生人数登记表,毕业生登记表,经费收支情况登记表,等等。

2. 专题性资料

如果要分析、研究某一专门的问题,仅靠经常性资料是远远不够的,必须进行专题性的教育调查或实验。通过专题性的调查或实验所获得的资料称为专题性资料。

(1) 教育调查

教育调查是指在没有预定因子、不施行控制的条件下,对现成的教育方面有关客观事实所进行的观察和分析。它是教育科学研究中普遍采用的一种方法。

从调查方法来分,可分为现情调查、回顾调查和追踪调查。现情调查是指对当前正在发生的或存在着的事物所进行的调查。如对中学教师学历的调查。回顾调查是指用追溯方法来探索造成结果的因素,从果到因地研究问题的方法。如对犯罪青少年的犯罪原因的调查。追踪调查是指对同一批调查对象,在较长时期内作间隔性的观察分析。如对某县100名妇女的识字量,在十年之内每年测试一次,以了解其复盲情况的调查。

从调查范围来分,可分为全面调查和非全面调查。全面调查就是在一定范围内的普查。如对某城市中学生所进行的视力普查。非全面调查是指对从总体中抽取的一部分对象所进行的调查。在非全面调查中又分为抽样调查和典型调查。如上例,限于时间和人力,不能对该城市所有中学生的视力一一进行测试,可从中随机抽取数所中学进行测试,以了解全城市的情况,这就是抽样调查。若从该城市中挑选视力保护工作较好的(或较差的)中学进行测试,以总结经验(或教训),这就是典型调查。

为了通过调查取得真实、准确、可靠、系统、完整的数据,必须有周密的调查计划,不仅应当明确调查目的、指标、方法,而且应当明确调查所获得数据的类型,以及统计处理的方法。

(2) 教育实验

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

为了突出实验因子的作用,排除非实验因子的干扰,显现实验者所操纵的自变量对因变量的影响,一般设立两种实验处理进行对照和比较。根据两种实验处理对照比较的形式不同,分为单组实验、等组实验和轮组实验。单组实验是指对同一组实验对象先后施行两种实验处理。如,先以每人单独的形式跑 100 米,休息一小时后,再以两人竞赛的形式跑 100 米,看两种赛跑形式对跑 100 米所需时间的影响。等组实验是指在甲、乙两组条件基本相同的情况下,对之施行不同的实验处理。为了使两组对象条件相同,也可以将两组对象依据条件相同的原则一一配对。如根据两组学生的年龄、年级、性别、近视情况一一配对,然后一组学生采用雾视法(即每日戴老花眼镜远眺半小时)治疗假性近视眼,另一组学生不戴眼镜或每日戴平光眼镜远眺半小时,治疗假性近视眼,以对照比较两种治疗方法的效果。轮组实验是指在实验组和对照组分别进行两种实验处理,并且每种处理各重复一次,也即每个或多个单组实验的联合。假如在甲、乙两组进行 A、B 两种实验处理,甲组实验次序为先 A 后 B,重复时则为先 B 后 A;乙组实验次序为先 B 后 A,重复时则为先 A 后 B。仍以跑 100 米的形式不同为例来说明轮组实验。甲组第一天先以每人单独形式跑 100 米,休息一小时后,再以两人竞赛形式跑 100 米。第二天,甲组先以两人竞赛形式跑 100 米,休息 1 小时后,再以每人单独形式跑 100 米。而乙组第一天先以两人竞赛形式跑 100 米,休息 1 小时后,再以每人单独形式跑 100 米。第二天,乙组先以每人单独形式跑 100 米,休息 1 小时后,再以两人竞赛形式跑 100 米。这种布置可以消除因时间顺序不同所带来的误差。

二、数据的种类

数据是随机变量的观察值。它是用来描述对客观事物观察测量结果的数值。当我们对某个随机变量进行观测时,事先不能预料会取到什么值,一旦某个值被取定,就称这个值为随机变量的一个观察值,即数据。

数据的种类不同,统计处理的方法也不同。统计数据按来源可分为点计数据和度量数据;按随机变量取值情况,可分为间断型随机变量的数据和连续型随机变量的数据。

1. 点计数据和度量数据

点计数据是指计算个数所获得的数据。如学校数、班级数、学生数、教师数、课程数、教室数、教学仪器数等。

度量数据是指用一定的工具或一定的标准测量所获得的数据。例如,用测高器测得学生身高的数据,用秒表测得学生完成某种作业所用时间的数据,用某种智力测验测得学生智商的数据,用某学科测验获得学生该科知识、能力、掌握情况的数据,等等。

2. 间断型随机变量的数据和连续型随机变量的数据

取值个数有限的数,称为间断型随机变量的数据。这种数据的单位是独立的,两个单位之间不能再划分成细小的单位,一般用整数表示。例如,三好学生人数、某门学科不及格人数;学生的智力、学科成绩等项指标按优劣程度分别排列的名次;用 1、2、3、4、5 五个等级,对学生的品德、兴趣、爱好、活动能力等所打的成绩,都属于间断型随机变量的数据。

取值个数无限的(不可数的)数据,称为连续型随机变量的数据。它们可能的取值范围

能连续充满某一个区间。数据的单位之间可以再划分成无限多个细小的单位。数据可以用小数表示。例如,学生的身高、体重、智商、用百分制分数表示的学科成绩、完成作业所用的时间等,都属于连续型随机变量的数据。

三、数据的统计分类

数据的统计分类,是指按照研究对象的本质特征,根据分析研究的目的、任务,以及统计分析时所用统计方法的可能性,将所获得的数据进行分组归类。分类时要抓住研究对象的本质特性,并对它的概念作出明确、严格的界定。分类标志(即分组所依据的特性)既要明确,又要前后一致,也能将全部数据包含在内。分类不仅以研究对象的本质特性为依据,还要以研究目的、任务的需要为依据,当然也应考虑到统计方法的可能性。

分类标志按形式划分,可分为性质类别和数量类别。性质类别是按事物的不同性质进行分类。这种分类不表明事物之间的差异。例如,将学生分成男生与女生,将实验对象分成实验组与对照组,将作文成绩分为甲、乙、丙、丁,将健康状况分为好、中、差等。虽然后两例的分类,有好坏之分,但不能比较相差的数量。性质类别还可以进一步分成不同的层次。如本章第二节的表 2.6 是按性别和操行评定等级两个层次进行分类;表 2.7 是按班级、性别、操行评定等级三个层次进行分类。数量类别是按数值大小进行分类,并排成顺序。在排列顺序时,可以直接按数值大小进行排列,也可以用等级顺序进行排列。

第二节 统计表

统计表是用来表达统计指标与被说明的事物之间数量关系的表格。它可以将大量数据的分类结果,清晰、概括、一目了然地表达出来,明显地反映出事物的全貌及其蕴含的特性,省去冗长的文字叙述,便于分析、比较、计算和记忆。

一、统计表的结构及其编制的原则和要求

统计表一般由标题、表号、标目、线条、数字、表注等项构成。

编制统计表的基本原则是:表的结构要简单明了。一张表只能有一个中心,说明的问题要重点突出,一目了然,避免绘制臃肿的包罗万象的大表;表的层次要清楚,项目、指标的排列要按照逻辑顺序合理安排。

统计表的基本格式如表 2.1 所示。现按表的组成部分来说明编制的要求。

表 2.1 统计表的基本格式

△△△△△△

(顶线)	
横标目的总标目 (亦可空白)	纵 标 目 (一般设谓语)
横 标 目 (一般设主语)	数 字
(底线)	

1. 标题

标题是表的名称,应确切地、简明扼要地说明表的内容。标题应写在表的上方。必要时,应在标题下注明资料的来源(地点、单位)和时间。

2. 表号

表号是表的序号。若文章中有几张表,则需按它们出现的先后次序编上序号,并写在标题的左方。

3. 标目

标目是表格中对统计数据分类的项目。按标目在表中的位置,可分为横标目和纵标目。位于表的左侧者为横标目,因为它与所指定的数字在同一横行;位于表的上端者为纵标目,因为它与所指定的数字在同一纵列。必要时,在横标目或纵标目的上方加上适当的总标目。按标目的内容又可以分为主语和谓语。主语是统计表叙述的对象,谓语是用以叙述的统计指标。由于横行书写的习惯,一般把主语放在横标目上,谓语放在纵标目上。如表 2.6 就是把主语(男、女学生)放在横标目上,谓语(甲、乙、丙、丁等级)放在纵标目上。但当主语较复杂时,可移动一部分到纵标目上去;如谓语较复杂时,可移动一部分到横标目上去。表 2.7 就是把主语的一部分(男、女学生)移到了纵标目上。一个设计良好的统计表,循“主语—谓语—数字”自左向右的顺序阅读,可以形成一个通顺的句子。例如表 2.6,可读为:某年级操行评定结果,男同学甲等 14 人,乙等有 30 人,丙等有 14 人,丁等有 4 人;女同学甲等 17 人,乙等有……

4. 线条

线条不宜过多。顶线、低线、隔开纵标目与数字的横线,以及隔开横标目与数字的纵线,是表的四种基本线条。如设有总和,可在总和上方添一横线。其余线条应尽量减少。特别是表的左上角不宜有斜线,表的中间不要有横线,表的左右两侧不要用纵线封闭。

5. 数字

表内数字必须准确,一律用阿拉伯字母表示,位次对齐,小数的位数一致。表内不应有空格。暂缺或未记录可用“…”或“……”表示,无数字用“—”表示,数字若是“0”,则应填写“0”。

6. 表注

它不是表的必要组成部分。若确有必须补充说明的问题,可用简短的小号字写在表的下方。

二、统计表的种类

1. 简单表

只列出观察对象的名称、地点、时序或统计指标名称的统计表为简单表。如表 2.2、2.3、2.4 所示。

表 2.2 某年级各班学生人数

班 别	一班	二班	三班	四班	总和
人 数	42	36	50	45	173

表 2.3 某校高三学生各年高考录取人数

年 份	1996	1997	1998	1999	2000	总和
高考录取人数	132	154	144	123	125	678

表 2.4 某班各科期末成绩

学科名称	语 文	数 学	英 语
平均分数(\bar{X})	71	80	78

2. 分组表

只按一个标志分组的统计表为分组表。如表 2.5 就是只按年龄这一标志分成各年龄组。

表 2.5 上海市区男幼儿 20 米跑步用时

年 龄 组	3 岁—	4 岁—	5 岁—	6 岁—
平均秒数(\bar{X})	7.71	7.16	6.04	5.53

资料来源:引自《华东师范大学学报》(教育科学版),1985 年第 2 期第 30 页。

3. 复合表

按两个或两个以上标志分组的统计表为复合表。如表 2.6 是按性别和操行评定等级两个标志进行分组的。按两个标志分组的统计表称二项表。表 2.7 是按照班级、性别、操行评定等级三个标志进行分组的。按三个标志分组的统计表称三项表。

表 2.6 某年级操行评定结果

性 别	甲		乙		丙		丁		总和
	人数	%	人数	%	人数	%	人数	%	
男	14	22.58	30	48.39	14	22.58	4	6.45	62
女	17	29.31	24	41.38	16	27.59	1	1.72	58
总 和	31		54		30		5		120
百分比	25.83%		45%		25%		4.17%		100%

表 2.7 某年级操行评定结果

班 别	甲		乙		丙		丁		总和
	男	女	男	女	男	女	男	女	
一 班	6	5	8	8	6	4	2	1	40
二 班	5	5	9	10	3	3	1	1	37
三 班	7	6	9	8	4	3	0	1	38
总 和	18	16	26	26	13	10	3	3	115

三、频数分布表列法

某一个随机事件在 n 次试验中出现的次数称为这个随机事件的频数。各种随机事件在 n 次试验中出现的次数分布称为频数分布。将其用表格形式表示出来称为频数分布表。例