

前 言

《教育统计与测量》含概教育统计学和教育测量学两门学科。教育测量主要研究如何编制科学的测验工具(教育测验、调查问卷等),阐述鉴定测验工具的可信性和有效性的理论和方法,以保证测验能搜集到可信、有效的资料。教育统计则研究如何整理数字资料,如何根据数理统计原理对资料进行统计分析做出推论,科学的揭示教育现象的特征和规律。教育统计和教育测量都是应用性极强的工具性学科。教育统计与测量作为测量工具和科研手段,在教育诊断、身心发展、智能评估、因材施教、人才选拔、职业指导,以及理论研究诸方面得到广泛的使用。

高科技发展的现代教育,也快速发展,当今的教育工作者,不仅要有丰富的科学知识,还必须具备科学研究的能力和技能,而教育统计与测量就是培养科研能力和技能的基本学科。所以,《教育统计与测量》是师范教育的必修课,掌握教育统计与测量教师和教育工作者的一项业务素质。1994年,辽宁师范大学在公共教育理论课的改革中,突破了只开设教育学、心理学两门教育理论课的传统模式,便增设了《教育统计与测量》课程。5年来的教学实践,已取得了良好的成效。

本教材在使用中,及时综合来自各方面意见,几经修改以使其更加完善。在编写中参阅了多种版本的统计学、测量学著作,选材及编写突出了实用性和可操作性。本教材不仅使用于师范院校本科教学,也可使用于成人教育专科和本科教学,亦可供广大中小学教师和教育行政管理者参考使用。

本书编写由魏华忠教授主持,周仁来博士、马健生博士参加三人合作编著。其中周仁来编写了第十四、十五、十六间,并参与第

三章第四节、第四章第一节、第五章第一节、第六章第一节和第七章的编写;马健生参与了第九章第二节(部分)第四节、第十章第二节部分、第三节、第三节,第十一章三、四部分,第十二章四、六(部分)以及第十三章第一节、第二节部分的编写。全书统稿和定稿由魏华忠负责。

本教材由辽宁省高等教育自学考试委员会组织编写,并得力于辽宁师范大学教务处、辽宁师范大学出版社以及我校教育系等各级领导的帮助与指导,在此致以深深的谢意。在编写中参阅和引用了许多专家的有关著作,在此一并致谢了。本书虽多次修订,今天再次出版,仍难免存有错漏,还请广大读者和同行专家批评指正!

编者

1999年3月于大连

目 录

第一章 绪论	1
第一节 教育统计学概述	1
第二节 统计学的几个基本概念	3
第二章 数据的整理	6
第一节 数据	6
第二节 数据的整理	8
第三章 描述统计	26
第一节 集中量数	26
第二节 差异量数	37
第三节 地位量数	45
第四节 相关量数	54
第四章 概率与分布	75
第一节 概率	75
第二节 分布	79
第五章 抽样分布与统计推论	96
第一节 抽样方法	96
第二节 抽样分布	98
第三节 总体参数估计	105
第六章 假设检验	116
第一节 假设检验的原理与步骤	116
第二节 样本与总体的差异检验	122
第三节 两总体均数量的差异显著性检验	127
第四节 其他统计量的差异显著性检验	136
第七章 方差分析与 F 检验	142
第一节 方差分析的基本原理	142
第二节 单因素完全随机化设计的方差分析	147
第三节 双因素完全随机化设计方差分析	156
第八章 χ^2 检验	165
第一节 拟合性检验	165
第二节 独立性检验	172
第九章 教育测量概述	182
第一节 关于教育测量	182

第二节	教育测验	188
第三节	教育测验的编制	201
第四节	良好的教育测验应具备的条件	213
第十章	测验题目的质量鉴定	218
第一节	测题的难度分析	218
第二节	测题的区分度分析	226
第三节	目标参照测验的测题分析	234
第四节	测题的分析程序	237
第十一章	测验的信度鉴定	242
第一节	测验的方法及其估计方法	242
第二节	信度的相关问题	254
第十二章	测验的效度鉴定	261
第一节	测验的效度	261
第二节	效度的应用	274
第十三章	测验分数的处理	286
第一节	处理测验分数的必要性	286
第二节	测验分数的转换	288
第三节	测验分数的解释	301
第四节	测验分数的合成	304
第十四章	智力测验	310
第一节	智力概述	310
第二节	个别智力测验	325
第三节	团体智力测验	337
第十五章	人格测验	345
第一节	人格测验的内容	345
第二节	投射测验	346
第三节	自陈量表	355
第十六章	潜能发展测验	367
第一节	性向区别测验	367
第二节	职业兴趣测验	369
参考文献		376
统计用表		378

第一章 绪论

第一节 教育统计学概述

一、教育统计简介

教育与心理教育统计学是应用统计学的一个分支,是数理统计学与教育学、心理学的一门交叉学科,它把统计学的理论方法应用于教育实际工作和各种心理实验、心理测验等科学研究中,通过对所得数据的分析和处理,达到更为准确地掌握情况、探索规律、制订方案、检查效率的目的,为教育和心理科学研究提供一种科学的研究方法。

1904年,美国心理学家 E.L.Thorndike 出版了《心理与社会测量导论》(An Introduction to the Theory of Mental and Social Measurements)一书。他在该书中试图以心理学和统计学为工具研究教育学,以使教育科学化,该书被看作是世界上第一本有关教育与心理统计学的专著。随后,美国一些大学先后开设了教育与心理统计学课程,有关的专著、教材也陆续出版,如 L.L. Thurstone 的《统计学纲要》(1924), H·E·Garrett 的《心理与教育统计法》(1926)。我国当时的部分院校也开设了这门课程,也出版了一批专著、教材。如薛鸿志的《教育统计法》(1925),周调阳的《教育统计学》,朱君毅的《教育统计学》(1930),王书林的《教育测验与统计》。这一时期的教育与心理统计学主要讲授描述统计学的内容。

四十年代后,教育与心理统计学转入以推断统计为主要内容的时期,小样本理论、参数估计、假设检验、非参数分析等内容被充实到教育与心理统计学中。一大批优秀著作也涌现出来,如 E·F·Lindquist 的《教育研究中的统计分析》(1940), W·J·Popham 和 K·

Sirotinik 合著的《教育统计学—应用与说明》(1967), J·P·Guilford 和 Benjamin Fruchte 合著的《心理与教育中的统计学》(1978)。国内也有一批著作出版, 如沈有乾的《教育统计学讲话》(1946), 《实验设计与统计方法》(1947)等。建国以后, 尤其是“文化大革命”期间, 由于众所周知的原因, 教育与心理统计学的研究和应用遇到了阻力。而同一时期, 国外的教育与心理统计学得到了长足发展。

70 年代以来, 由于计算机技术、信息科学的迅速发展和应用, 使得短时间内进行大量的复杂运算成为可能, 同时, 人们已不满足于复杂现象的简单分析和预测, 多元统计分析也在教育和心理研究中得到应用。各种专用统计软件使得数据处理能够快速、准确、便利, 把心理与教育统计学推向更为广泛的应用领域。自 1979 年我国恢复教育与心理统计学的教育和研究以来, 一批高质量的优秀学术著作和教材得以出版, 如叶佩华的《教育统计学》, (1983), 张厚粲的《心理与教育统计》(1982, 1987), 左任侠的《教育与心理统计学》(1982), 郝德元的《教育与心理统计》(1982), 以及许祖慰的《项目反应理论及其在测验中的应用》(1992), 张敏强的《教育与心理统计学》(1993), 王权的《现代因素分析》(1993), 舒华的《心理与教育研究中的多因素实验设计》(1994), 阮桂海的《微电脑的使用与软件应用大全——windows 之 spss6.0 高级应用教程》(1995)等。

二、教育统计学的内容

教育与心理统计学的内容和应用领域继续在不断扩展, 在提供决策信息和实验结果分析方面的作用更加明显。在内容方面更加朝实用性方向发展, 既强调统计分析方法的应用范围和应用条件, 也强调计算方法与计算技术的紧密结合。在实验方面, 更重视各种实验设计方法, 重视数据的来源, 强调定性定量研究的有机结合, 从而使得对数据的分析、处理更具有科学性和预见性。

教育与心理统计学的内容主要分为三个方面:

1. 描述统计学。主要讲述对从大量观测中获得的数据进行整理,以描述事物的典型性、波动范围以及相互关系,揭示事物背后所隐藏的内部规律。包括次数分布表与图的制作和各种统计量的计算,如集中量、地位量、差异量、相关量等等。

2. 推论统计学。主要讲述如何由部分数据去推断整体的方法。包括统计检验、统计分析和非参数统计等内容。

3. 多元统计分析法。主要讲述对超过两个因素的教育与心理实验和研究的统计方法。主要包括主成分分析、因素分析、聚类分析、多元方差分析、多元回归分析等。

第二节 统计学的几个基本概念

一、随机现象与随机变量

随机现象是指,在一定条件下进行的实验或观察,某一事物发生的结果不止一个,哪种结果发生事先不能确定,具有偶然性,但经过反复多次的观察,结果的发生呈现一定的规律性,这类现象称为随机现象。比如掷一枚硬币,其结果有两种:或是正面,或是反面,事先无法确定,但对发生正面多次观察,便会发现正面(或反面)发生次数约占总数的 $1/2$,具有规律性。掷一枚硬币,发生正面或反面的现象便是随机现象,正面的发生或反面发生叫随机事件。

随机事件发生的结果各次不尽相同,记载它的量受随机因素影响,具有随机性,其特点既有偶然性,又有规律性,我们将它称为随机变量。随机变量可表为一列、一组或一堆观测值,统计学中又叫数据或数据资料。随机用大写的 $X_i, Y_i, Z_i \cdots$ 等表示, X_i 即 $X_1, X_2, X_3 \cdots X_i, X_n$ 。

二、总体、个体、样本

任何一项研究都有其研究对象,那么,研究便会遇到总体个体

和样本的问题。

(1)总体 总体即观测对象的全部个体。总体的界定,随研究问题而变动,如若研究“中国小学生的学习能力”,总体则指中国所有小学生;你没有能力做到对这么大的总体进行研究,可以提出研究“ $\times \times$ 市小学生的学习能力”,这时,总体仅指 $\times \times$ 市所有小学生,所以在具体研究中,总体是相对的。

依据总体的个体数量而言,总体有无限总体和有限总体;依据个体的性质而言,总体又有抽象总体与具体总体之分。

(2)个体 个体即构成总体的单位。一种个体指具体的人或物,一个人、一个物就是一个个体,其构成的总体是具体总体;一种个体指对某种现象如智力、学习能力等的一次观测结果,观测 10 次,就得到 10 个个体。这种个体属于理论的,抽象的个体,其总体是抽象总体,往往是无限的。统计中个体数一般都记作次数(f)。

(3)样本 样本是由总体的一部分构成的,样本必须对总体具有代表性,不能代表总体的样本便不成其为样本。样本代表性取决于样本正确取样和足够的容量(n)。样本容量就是样本中的个体数目多少,在教育统计学中,样本容量 $n \geq 30$ 为大样本, $n < 30$ 的样本上为小样本。所谓正确取样就是,取样时必须遵循随机性,即坚持随机取样原则。随机取样的方法有简单随机取样法,如抓阄法,随机表达法,分层随机取样法等。

三、次数、频率、概率

(1)次数是指某一事件发生的数目,通常用符号“ f ”表示。例如一个班级通过某测验时,成绩为 80 分的共有 10 个,10 就是 80 分这一事件出现的次数,即 $f = 10$ 。

(2)频率是指在一组数据中,某一事件(某一数值)出现的相对次数。即某事件出现的次数被一组数据的总次数除的比例。

(3)概率是无限的观测某一事件在总体中出现的频率稳定比值。假设频率为 p ,总次数为 n ,某事件出现的次数为 m ,则 $p = m/n$

n。这是在有限总体中某件的频率,若作无限次的观测,该事件频率趋于稳定在一个比值上。频率的稳定值就是该事件发生的概率。概率只是个理论值,一般用频率作为概率的代表值。

四、误差

误差通常指观测值与真值之差。在观测中由于种种条件的限制和影响,误差难免产生,误差大致有以下几种:

(1)随机误差:亦称偶然误差,它是由一些偶然因素引起的,使观测值与真值发生误差。随机误差是绝对存在不可避免的,但随着观测次数的增加,随机误差的变化会呈现一定的规律性,即它总是围绕真值波动,有增有减,因此,偶然误差在大量观测中可趋于抵消。

(2)系统误差:又称恒定误差,是观测中服从确定性规律的误差。例如:同一教师分别采取两种教学方法,采取优等方法的学生考核成绩偏高,这样对照于采取普通方法的学生便呈现出恒定的偏高误差,这类误差不能够通过增加实验次数消除。只能通过一定的方法才能消除。

(3)过失误差:由于实验观测过程中的过失造成的误差。如操作失误,仪表失灵、遗漏、部分丢失等。过失误差是不允许的,在数据整理中若发现有过失误差的数据(称为异常点),应删除,以保证计算结果的准确性。

这三种误差总是错综复杂地存在着。因此在实验中有误差是难免的,有些误差是允许的,有些误差是不允许的。

5. 统计量与参数

在科研实验中,描述样本情况的一些统计指标称为统计量,如平均数、标准差、相关系数等。而描述一个总体情况的一些统计指标称为参数。如总体均数、总体标准差、总体相关系数等。当参数不易获得时,在某种程度上可用样本的统计量作为总体参数的估计值。统计量与参数使用符号不同,学习时应注意区别。

第二章 数据的整理

第一节 数 据

一、数据特点

在教育与心理测验、实验等的研究中,会得到大量的数据资料。在对这些数据进行初步的整理之前,首先要了解数据所具有的一些特点。

第一,离散性。随机事件发生的结果具有偶然性,前一次结果与后一次结果没有必然联系,即每次结果之间都无必然联系。所以,每次结果的数量化都是独立的,互不关联的数据。例如,考察六岁儿童身高,测得了一百个儿童的身高会得到一百个数据,这些数据之间是相互独立的,无必然联系,这就是数据的离散性。

第二,波动性。又叫变异性,随机事件的发生有多种可能性,每次观察得到的测验值都可能是不一样的,有时大,有时小。测量仪器越是精密,测得的数值差异越容易显现出来。仍以测量6岁儿童身高为例,谁高谁低是随机的,得到的大量观测值也必然表现出波动性。当然,这种波动性是在一定的时空范围内产生的。

第三,规律性。大量观察、实验得到的数据虽然呈现一定的波动性,但在一定范围内,趋向于稳定在某一数值上,从而呈现出一定的规律性。正如上例所述,6岁儿童身高的波动总趋向于稳定在某一数值上,显示出6岁儿童身高的特征。

二、数据的种类

在对数据进行统计处理前,要对数据先进行分类,根据不同类型数据的性质、特点进行不同的统计处理。对数据进行分类,通常从两个方面加以考虑。

1. 按数据取得的方法的不同,将数据划分为计数数据和计量数据。

(1)计数数据,是计量个数(或次数)的数据。例如,学生人数、人口数、学校数、班级数等。它具有独立的,不可分割的单位,所以,一般采取整数形式如1个,1次等。

(2)计量数据,是借助于一定的测量工具或一定的测量标准得到的数据。这类数据依测量工具或测量准则的不同,可分为与类别测量、等级测量、等距测量、等比测量相应的四种数据。

2. 根据数据是否具有连续性,将数据划分为离散变量和连续变量。

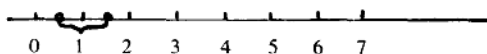
(1)离散变量,其数据的计量单位都是整个个体,不能再分割,而且数据之间无连续性,也称为间断变量。它在数轴上是一个点,点与点之间没有连续性,在数轴上记为:



(2)连续变量,其数据的计量单位可以分割,单位之间是不间断的连续的。它在数轴上是一段距离,而非一个点,记为:



在量尺上记为:



这时的1是指从0.5~1.5之间的一个单位距离,连续变量1、2、3……经常被表示为0.5~1.5,1.5~2.5,2.5~3.5,……的形式。这是因为,在实际观测时,由于误差的原因,作为连续变量的观测数据只代表一个单位的中央点,对它的表示法、理解及计算都有所不同。因此,在对数据进行整理时,要判断数据的种类,尤其要辨别连续变量与间断变量,并据此进行不同的统计处理。

第二节 数据的整理

一、次数分布表

次数分布的表和图是数据整理和描述的基本形式。编制次数分布表、绘制次数分布图,能够了解一组数据的分布情况。尽管表和图的绘制我们并不陌生,生活中也司空见惯,它们在数理统计中却占有相当重要的地位。

1. 次数分布表

次数分布表是通过将散乱的观测数据,划分为若干组,以数据发生的次数归入组内列成次数分布表,使数据井然有序,隐含于数据中的研究内容的特征得以显现出来,同时也能显示出数据分布的理论性质。它也是进行统计处理的依据。

下面根据一次测验成绩获得的数据资料说明次数分布表的编制。

25	35	37	55	27	40	33	39	51	28
42	27	34	44	29	36	22	51	29	21
28	53	29	40	43	33	33	42	36	41
20	25	38	47	46	16	32	36	13	37
27	33	46	10	31	34	32	32	43	14
30	46	20	21	37	19	24	24	21	33

(1)求全距(R)。全距指一组数据中最大数与最小数之差,它表示一组数据的最大波动范围。符号为 R。即,

$R = \text{数据最大值} - \text{数据最小值}$

上例资料中,60 个数据的最大值为 55,最小值为 10,所以,

$R = 55 - 10 = 45$

(2)确定组距与组数。组距是每一组的间距,即每一组包含的单位,用 i 表示。常采用的组距为 2、3、5、10、20 等。组数根据数据的多少和数据的性质而定,用 K 表示。分组多,组距小,误差相

应减少,组计算繁杂,组距太小,规律性不能突出,失去分组意义。分组少,组距大,误差也随之增大,会隐藏数据的特点。组距大小、组数的多少应视具体情况来定。通常,若数据的个数在 200 个以内,一般分为 8~18 组。如果数据来自一个正态的总体,可利用下述经验公式确定组数,即

$$K = 1.87(N - 1)^{\frac{2}{5}}$$

上述公式中的 N 为数据个数,就上例 60 个数据而言, $N = 60$, 按此公式计算后取整, $K = 10$ 。当然,在实际分组时,要考虑数据的具体情况而与计划分组数略有不同,这是正常的。上例中,组距 $i = 5$,组数 $K = 10$ 。

(3) 列出组限。组限是每组的两个极限,其最低数为下限(L_b),最高数为上限(L_a)。

上例中, $i = 5$,组数从最小数开始,即 10~15 为第一组,15~20 为第二组,20~25 为第三组,依此类推。但是,对于组限,有几点需要加以注意:

①组限的写法。下述五种是组限的常用写法(见表 2—1),表达同一含义,可自行选择。

表 2—1 组限的常用写法

(1)	(2)	(3)	(4)	(5)
15~	15~19	15~20	14.5~19.5	17
10~	10~14	10~15	9.5~14.5	12

②组限的理解。次数分布表处理的是连续变量,10~15,15~20 在数轴上是连续的,不间断的。因此,10~15 应理解为是 10 分到 14.999……分,即不到 15 分,而 15 分已属于 15~20 一组了。

③运算时的组限。由于每一组的上限都是一个极限(如 14.999…),无法进行计算,所以,必须确定每组上下限的精确值。如表 2—1,第一行的精确组限都是 14.5~19.5(不包括 19.5 本

身。这里要指出的是组中值(X_c),即每组上、下限的中点值,通常用它来代表每组的测验值。

$$X_c = \frac{\text{精确下限} + \text{精确上限}}{2}$$

(4)登记次数。采取类似唱票的办法,统计落在每一组限范围内的数据次数。通常采用记“正”的形式,这一工作量可在草纸上进行,核实无误后再写到正式的次数分布表上。

(5)整理成正式次数分布表。次数分布表的主要项目是分组和次数。同时依据统计处理的需要,还可以有组中值、频率、上累加次数(即下限以上次数的累加)以及比例、下累加次数(即上限以下次数的累加)及比例、离差等。

根据上述所列的五个步骤,将前面的 60 个数据整理制成次数分布表如下(表 2—2)。

表 2—2 一次测验的次数分布表

组数	组距	组中值	次数	频率	下累加次数、比例		上累加次数、比例	
K=10	i=5	X_c	f	P	f_b	P_b	f_a	P_a
1	55-59	57	1	0.017	60	1.00	1	0.02
2	50-54	52	3	0.050	59	0.98	4	0.07
3	45-49	47	4	0.067	56	0.93	8	0.13
4	40-44	42	8	0.133	52	0.87	16	0.27
5	35-39	37	9	0.150	44	0.73	25	0.42
6	30-34	32	13	0.216	35	0.58	38	0.63
7	25-29	27	10	0.167	22	0.37	48	0.80
8	20-24	22	7	0.117	12	0.20	55	0.92
9	15-19	17	3	0.050	5	0.08	58	0.97
10	10-14	12	2	0.033	2	0.03	60	1.00
			$\Sigma f = 60$					

从表 2—2 中能够看出,制作次数分布表的意义。将搜集来的大量原始数据整理成次数分布表的形式,能直观地显示出数据的分布情况,如哪一组次数多,数据大多集中在哪,体现了数据的集

中趋势;随着次数的减少、增加,体现了数据的离散情况,描述了一系列数据的“地形”分布。表中列出的频率、累加次数对以后进一步分析、研究各组概率的发生提供了方便,还有助于理论次数分布的研究,然而,次数分布表也有其不足。数据归入组内,以组中值作为每一组的代表值,而这是假定次数在一组内是均匀分布的,大于或小于组中值被当作是这些数据的平均数,从而使得计算中不可避免地引入了误差。所以,在实际计算中,为减少误差,应处理好一组数据的组距及分组。

二、次数分布图

有了次数分布表,对数据分布的全貌就有了了解。但将次数分布以图形的方式显示出来,数据的分布则更显示出直观性。常用的次数分布图有直方图和多边图。次数分布图是依据次数分布表绘制的,下面依据表 2—2 作次数分布图(见图 2—1 至图 2—3)。

1. 在横坐标上标出等距分组点作为分数,不一定从零点开始;纵坐标为次数,从零点开始。

2. 直方图相邻各直条间不留空隙,不能间断。

3. 多边图是将分组区间中点(组中值)相连法绘制成多边形。多边形中曲线的两端不与横坐标相接,可用虚线连接。

多边图与直方图都是以面积表示连续性数据的次数分布。多边图下的面积也为总次数,如果样本很大,能够描绘出一条分布曲线,这对总体的理论次数分布的分析很重要。直方图中,它的每一个直方条面积为该分数组的次数,直方条的面积总和为数据的总次数;每一个直方条的面积与总面积的比,为该直方条的分数发生的频率。

根据累加次数分布表还可以作累加次数分布图(见表 2—3)。累加次数分布图又有直方图和累加曲线图之分。不同的是,累加曲线连结的不是直方上端的中点,而是右侧的端点,因为它表示的

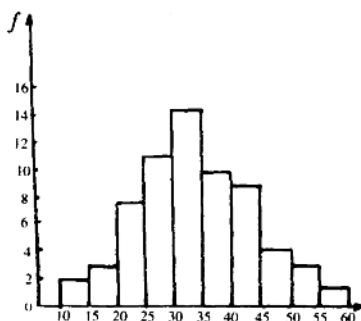


图 2—1 直方图

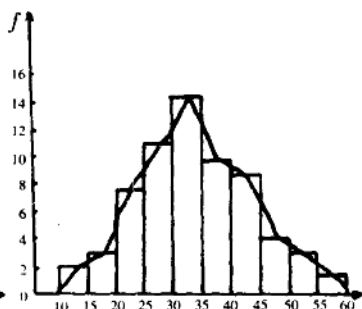


图 2—2 多边形

每组上限以下有多少次数,故不能用中点(组中值)来表示。

练习与思考题

1. 教育与心理统计学包括哪些研究内容?
2. 简述数据的特点。
3. 用下列数据作次数分布表和图(包括累加次数表和图)

72	58	67	68	58	56	62	29	91	78	52	48
57	73	77	59	59	48	62	51	39	91	57	43
68	58	63	86	70	55	77	62	58	58	30	42
80	53	44	63	52	77	48	60	79	81	43	85
72	46	63	58	54	63	67	51	51	72	59	73
56	53	81	53	64	63	76	67	67	58	58	82
56	56	69	52	88	48	60	68	68	29	53	43
69	44	88	63	55	50	88	51	51	55	51	

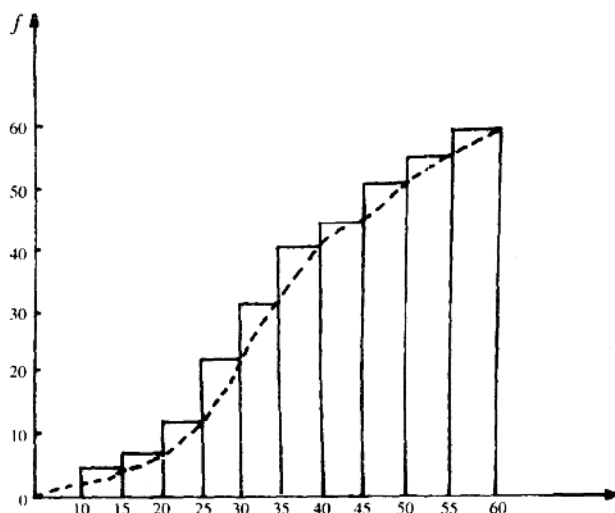


图 2—3 累加直方图和累加曲线图

附录一

一、统计表*

统计表是表达数字资料的一种重要方式,在数据进行统计分类以后,往往采用统计表的方式来表达。对数据进行分类后得到的各种数量结果称为统计指标。所谓统计表就是把统计指标和被说明的事物用表格的形式表示出来。换句话说,统计表是由统计指标和被说明的事物两部分组成的。统计表给人以一目了然、清晰简洁的印象,同时,在一个表中可以进行复杂的分类,数据易于比较。在教学、心理学的科研成果报告中,在宣传及展览中,统计

* 附录二选自张厚粲、孟庆茂编《心理与教育统计》(内部发行本)略做组编。