

教育统计入门

刘国铃 冯国卫



● 科学技术文献出版社重庆分社

教育统计入门

刘国铃 冯国卫

科学技术文献出版社重庆分社

教育统计入门

刘国铃 冯国卫

科学技术文献出版社重庆分社 出版

重庆市市中区胜利路132号

新华书店重庆发行所 发行

科学技术文献出版社重庆分社印刷厂 印刷

开本: 787×1092毫米1/32

印张: 6.5 字数: 14万

1987年6月第一版

1987年6月第一次印刷

印数: 10000

统一书号: 4176·27

定价: 1.30元

前 言

随着普教事业的深入发展和教育科研及教育改革的广泛进行,如何正确分析教育活动中搜集来的数据,并从中得出统计规律,从而科学地指导教育实践,已是广大教育工作者深感重要的事。为了适应这种形势的需要,我们编写了这本带入门性质的书。

本书是编者在多次讲授过程中广泛征求意见,经补充、修改而成。书中选入了当前普通教育事业中最需要、最实用的教育统计知识;摒弃了由于时代的发展相对显得陈旧的内容。在尽量避免罗列过繁的数学推导过程的同时,力求用通俗的语言和切合中、小学实际的例子讲清统计原理,并注意理论的严密性。为了提高运用教育统计知识的能力,各章节配有练习题,书末附有参考答案。

本书初稿在使用过程中,普遍反映具有深入浅出、通俗实用、易学易懂的特点,一般具有初中以上文化者就可读懂。

本书适合广大中小学教师、教研人员、教育行政干部阅读,也可供师范学校、教师进修院校师生学习参考。

由于我们水平有限,错误在所难免,欢迎广大读者批评指正。

编者

1986年10月

绪 论

教育统计学是研究如何将统计方法，特别是数理统计方法应用于教育问题的一门学科。

教育统计学可分为描述统计与推断统计两大类。

描述统计着眼于对所关心的教育现象进行全面调查或观察，从杂乱无章的资料中获取各种数据，并计算出各种统计数字，绘制各种统计图表，以此对总体作出有指导意义的结论。同时，它也是推断统计的基础。

推断统计的主要功能是从较少的一个群体（样本）的各种数据中获取信息，并把它合理地推广应用于较多的一个群体（总体）中去，从而得出科学的结论，用以指导教育实践和教育科学研究。

目 录

绪 论

第一章 描述统计.....(1)

第一节 统计数据的表列与图示.....(1)

第二节 算术平均数.....(9)

第三节 方差 标准差.....(19)

第四节 标准分.....(32)

第五节 变异系数.....(42)

第六节 相关系数.....(47)

第七节 利用计算器求相关系数.....(57)

第八节 信度 效度 难度 区分度.....(61)

第二章 推断统计.....(78)

第一节 概率基础知识.....(78)

第二节 总体与样本.....(86)

第三节 二项分布.....(94)

第四节 正态分布.....(101)

第五节 总体平均数的估计.....(109)

第六节 总体标准差的估计.....(117)

第七节 假设检验.....(121)

第八节 平均数的显著性检验.....(125)

第九节 方差的显著性检验.....(135)

第十节 比率的显著性检验.....(139)

第十一节 相关系数的显著性检验.....(146)

第十二节 正态拟合检验.....(151)

第十三节 一元线性回归.....	(157)
第十四节 一元方差分析.....	(161)
附：练习参考答案.....	(170)
附表 1 采用弗拉南根方法估计二列相关系数 γ_b 的数 据表.....	(174)
附表 2 正态分布表.....	(175)
附表 3 t 分布的双侧分位数 (α) 表.....	(180)
附表 4 χ^2 分布的上侧分位数 (χ_a^2) 表.....	(182)
附表 5 F 检验的临界值 ($F\alpha$) 表.....	(184)
附表 6 检验相关系数 $\rho = 0$ 的临界值 ($r\alpha$) 表.....	(196)
附表 7 r 与 Z 的换算表.....	(197)
附表 8 随机数表.....	(199)

第一章 描述统计

第一节 统计数据的表列与图示

在教育科学研究和教改实验中，要对所研究的对象进行调查和观察。调查和观察的结果量化后以数字形式表示出来，这些数字叫数据，如学生的成绩分数、身高、体重等等。

一群数据未加分类整理，显得杂乱无章，很难看出其中的规律性，不利于研究教育的对象。因此，整理搜集来的数据，使之系统化，就是教育统计的一项最基本工作。对数据的整理分两个方法：列表法和图示法。

1. 列表法

把数据进行归类汇总后，用表格的形式表现出来，这种表格叫统计表。统计表避免了冗长的文字叙述，表现数据系统、醒目，便于计算，容易检查遗漏和错误。

从格式看，统计表包括总标题、横行与纵栏标题、横行、纵栏、数据等要素。总标题写在表的上端中央，横行与纵栏标题写在上方和左方，数据写在右下方。

统计表有各种分类法，较普遍的是根据研究的对象指标类别不同而分为单项表、二项表、三项表等。

单项表:

表1-1 四川省1985年高考文科各科目试题平均难度表

科 目	政治	语文	数学	历史	地理	外语
平均难度	0.75	0.49	0.56	0.67	0.62	0.58

表1-1是根据平均难度一个指标分成不同科目的6个水平列的单项统计表。

二项表:

表1-2 一个假设的小学自然常识的布鲁姆知识与能力双向细目表

能力 知识	记忆	理解	应用	分析	综合	评价	合计
生物世界	3	5	6	3	2	1	20
资源利用	2	3	3	1	1	0	10
动力和机械	2	3	4	2	0	1	12
物质、物性与能量	5	6	8	3	2	1	25
气象	2	4	3	2	2	0	13
宇宙	2	5	4	1	0	0	12
地球	2	2	2	1	1	0	8
合计	18	28	30	13	8	3	100

表1-2是根据知识与能力两个指标列的二项表。

三项表:

表1-3 某教育学院某学期各专业各年级男女学员人数

系 别		中 文		外 语		数 学			总 计	
人 数	性别	男	女	男	女	男	女		男	女
	级别									
	一年级									
	二年级									
	总 计									

表1-3是根据系别、性别、级别三项指标列的统计表。

指标多于三个，制成一张统计表已失去醒目的意义，宜制成分类表。

2. 图示法

利用几何方法表示统计数据的图形叫统计图。统计图形象直观，通俗易懂，一目了然，能粗略看出数据之间的变化。

从格式看，统计图有总标题、坐标系、原点（零点）、比例、图例、单位、实线、虚线等等。总标题一般写在统计图的下边。

统计图种类甚多，常用的有条形图、扇形图、曲线图等。

条形图 用条形的高低或横向的长短来比较统计数据的几何图形叫条形图。条形图又分简单条形图和复合条形图。

简单条形图：

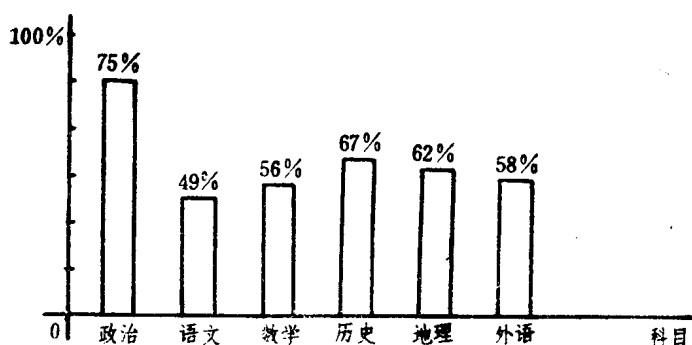


图1-1 四川省1985年高考文科各科目试题 平均难度统计图
复合条形图:

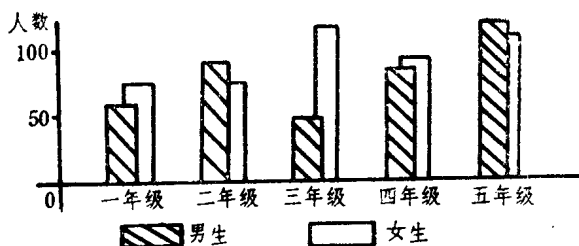


图1-2 某小学男女生人数情况

扇形图 把圆按比例分成若干扇形来说明总体结构的几何图形。

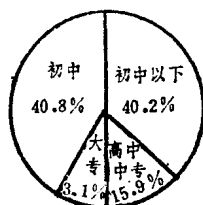


图1-3 1980年我国26省市
自治区2000万职工
文化水平统计

曲线图 用曲线的高低和斜度表现统计数据的几何图形叫曲线图（又叫图象）。

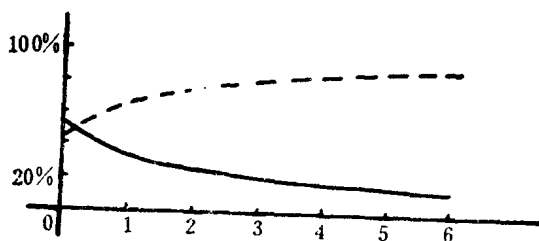


图1-4 艾宾浩斯无意义音节的保持和遗忘曲线

次数分布

次数分布是指一群数据中每一数据的数值（或一组数值在相应区间）所出现的次数的分布情况。把这种分布情况制成的统计图表分别叫做次数分布表和次数分布图。

次数分布表和次数分布图是教育统计中常用的和重要的图表。

次数分布表和次数分布图的编制一般按以下步骤进行：求全距，定组数，定组距，写组限，定组中点值，归类划记，登记次数（或相对次数、累积次数），列表，绘制次数直方图、次数多边形（次数曲线图）、相对次数直方图与多边形、累积次数曲线图等。

例1 某中学某班一次数学考试成绩表（见表1-4）。

1.求全距 一组数据最大值与最小值的差叫全距，用

表1-4

65	73	75	77	95	67	80	73	79	68
73	82	55	76	81	60	53	78	87	72
74	69	84	76	62	91	69	61	68	69
55	67	67	73	86	50	56	74	58	54
86	61	59	66	59	57	64	61	67	56

R表示:

$$R = \text{最大值} - \text{最小值} = 95 - 50 = 45$$

2. 定组数 组数要适中, 分得过多, 就失去分组和化繁为简的意义; 分得过少, 在近似计算中将引起较大失真。本例分成10组为宜。

3. 定组距 组数确定以后, 取全距与组数的商的过剩近似值作为组距。本例中, $45 \div 10 = 4.5$, 取5作为组距。

4. 写组限 组限即各组的分点。数值小的分点叫下限, 数值大的分点叫上限。确定分点的方法很多, 这里举一种常用而简捷的方法: 为了避免数据的值落在分点上, 我们把全距按组距的长度连续等分为若干个左闭右开的小区间 (用符号[)表示), 最后一个为双闭区间 (用符号[]表示)。分点就包括在“闭”的那个区间里。本例中组限就为50, 55, 60, …… , 100。

5. 定组中点值 为上下限值的和的二分之一, 常用 x_c

表示。本例中组中点值为 $\frac{50+55}{2} = 52.5, 57.5, \dots, 97.5$ 。

可把组中点值作为该组数据的代表值。

表1-5 某中学某班一次数学考试成绩分布表

组 别	组中点	划 记	次数 (个)	相对次数	累积次数	累积相 对次数
[50,55)	52		2	0.04	2	0.04
[55,60)	57	册	8	0.16	10	0.20
[60,65)	62	册	6	0.12	16	0.32
[65,70)	67	册册	12	0.24	28	0.56
[70,75)	72	册	7	0.14	35	0.70
[75,80)	77	册	6	0.12	41	0.82
[80,85)	82		4	0.08	45	0.90
[85,90)	87		3	0.06	48	0.96
[90,95)	92		1	0.02	49	0.98
[95,100]	97		1	0.02	50	1.00
总 计			N = 50	1.00		

6. 用类似唱票的办法把数据逐个划记入适当的组内。划记常用“册”符号或写“正”字。

7. 登记次数 各组划记的总数就是各组数据在相应组出现的次数，常用字母 f 表示。根据次数分布图表的要求，更常用的是把各组次数 f 与数据总个数 N 的商，即 f/N 作为相对次数填在表内。如果把累积次数（即各组次数按组数序号

逐个递加的各个和)或相对累积次数(即各组相对次数按组序号逐个递加的各个和)也填在表内,还可以分别得到累积次数分布表或累积相对次数分布表。表1-5即是一张综合的次数分布表。

根据次数分布表按统计条形图的制作步骤可以绘制次数分布直方图:建立坐标系,横轴表示数据的数值,纵轴表示次数,各条形高度为各组次数。见图1-5。

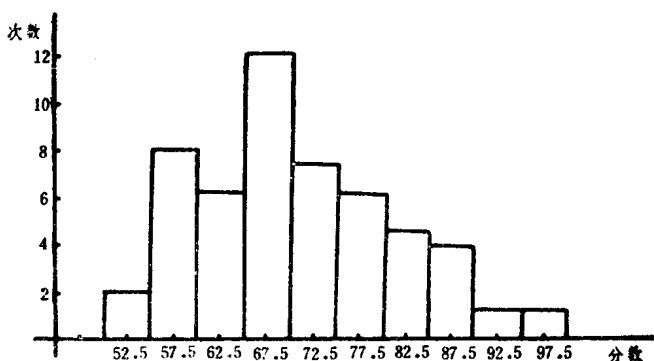


图1-5 某中学某班一次数学考试成绩次数分布直方图

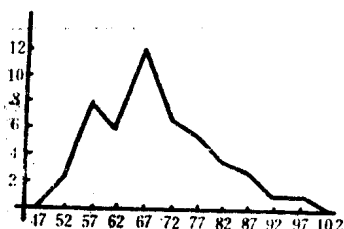


图1-6 某中学某班一次数学考试成绩次数多边形

如果把以各组中点值为横坐标、相应的次数为纵坐标得出的各点,以及横轴上两边各增加一个组中点,用折线顺次连起来,则可以得到次数多边形(见图1-6)。

下面再介绍一下累积次数曲线图。

如果把以各组下限为横坐标、下限以前相应的各组累积

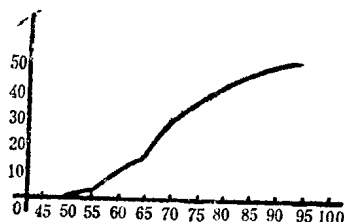


图1-7 某中学某班一次数学考试成绩累积次数曲线图

次数为纵坐标得到的各点顺次用平滑曲线连接起来，则可得累积次数曲线图（本例的累积次数曲线图见图1-7）。也可以从相反的方向累积并制图。

练习：

某校一年级两个班语文成绩如下：

67	62	83	51	83	59	79	81	93	78	46	50	97
69	65	44	74	74	43	64	73	90	94	44	91	62
96	59	78	43	79	64	71	74	95	43	73	56	76
76	73	54	66	76	69	64	71	73	74	68	83	79
85	68	66	61	74	67	62	99	72	62	79	79	55
77	63	75	71	69	78	68	82	84	95	54	72	66
79	81	72	71	54	77	74	68	76	72	90	50	55
68	84	82	81	81	54	73	77	80	73	94	58	84
69	82	51	70	69	74	83	90	75	78	72	68	63
68	45	96										

试绘制次数分布表和次数分布直方图。

第二节 算术平均数

对杂乱无章的数据进行初步整理后，可以得到统计图、表等，这种对数据粗略的、直观的概括是很有用的。但要进

一步深入分析研究，只有图表就不够了，还必须通过数据，计算出一些量数，用以说明数据的全貌、集中趋势、离散程度、分布特征、相关特征等各项特点。

描述分布中大量数据向某点集中情况的量数，叫集中量数，又称为数据的中心位置或集中趋势等。它包括：算术平均数、加权平均数、中数、众数、几何平均数、调和平均数等。本节我们重点介绍最常用的算术平均数和加权平均数。

集中量数的作用主要有两个，一是作为一组数据的典型水平或代表值，能说明这组数据全貌的一个方面的特征；二是可以通过集中量数对几组数据进行比较，判明彼此之间的数值差别。

1. 算术平均数

算术平均数就是各次观测所得数据的总和除以数据个数所得的商，通常简称为平均数或均数。用 $\bar{x}$ 表示变量 x 的平均数。

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{n} = \frac{\sum x_i}{n} \quad (2-1)$$

式中 $x_1, x_2, \dots, x_N$ 为 x 的 n 个观测数值，
 n ：观测次数。

(1) 由原始数据直接求平均数

由公式2—1可自原始数据直接求平均数。

例1 某学生的考试成绩为：政治85分，语文82分，数学94分，物理88分，化学76分，外语97分，则该生六科平均成绩

$$\bar{x} = \frac{\sum x}{n} = \frac{85 + 82 + 94 + 88 + 76 + 97}{6} = 87(\text{分})$$

(2) 由分组数据求平均数