教育评价的技术与方法

Techniques and Methods for Education Appraisal

主 编 向德全

西北大学出版社
NORTHWEST UNIVERSITY PRESS

教育评价的技术与方法

主　编：向德全

副主编：胡玲翠

　　　　宋　浩

西北大学出版社

图书在版编目（CIP）数据

教育评价的技术与方法 / 向德全主编. —西安：西北大学出版社，2006.10
ISBN 7-5604-2241-1

Ⅰ. 教…　Ⅱ. 向…　Ⅲ. 教育评估-研究
Ⅳ.G449

中国版本图书馆CIP数据核字（2006）第120940号

教育评价的技术与方法

向德全　主编

西北大学出版社出版发行

（西北大学校内　邮编710069　电话88302590）

新华书店经销　　陕西威思特印务有限责任公司

850毫米×1168毫米　1/32开本　11印张　286千字

2006年10月第1版　　2006年10月第1次印刷

ISBN 7-5604-2241-1/G·322　　定价：25.00元

前言

在科学研究活动中，要得出定量的结论，必须运用数学语言。马克思指出："一种科学只有在成功地运用数学时，才算达到了真正完善的地步。"在现在，数学分析方法已广泛地应用到各门学科的科学研究之中，不仅在自然科学，而且在社会科学、思维科学中都已普遍使用数学，产生了计量社会学、计量历史学、教育统计学、教育测量学等新的计量科学。教育评价学的研究同样离不开计量化的处理，同时也需要作数学分析处理。

科学研究的计量化过程，经历过三个主要发展阶段，即从精确数学到随机数学，再到现代的模糊数学。经典的精确数学，如数学分析、微分方程，它是用于研究必然现象或确定性的关系，主要用于自然科学领域。随机数学，如随机过程理论、数理统计，它是用来研究偶然现象，从纷乱的、大量的偶然现象中去探索必然的规律，在社会科学领域中得到广泛的应用。模糊数学，如模糊集合论，它是用来研究非精确现象，现在广泛地应用于社会科学和思维科学领域。

教育评价学研究的对象是教育教学过程和资源，包含大量的偶然现象和非精确现象。因此，要深入研究教育评价现象及其规律，必须运用统计描述、统计分析方法和模糊数学分析方法，才可能使这门学科达到真正完善的地步。也就是说，在构成教育评价学的诸多学科中，数学应该也必须扮演重要的角色。在教育评价过程中，必须自始自终贯穿数学的思想方法，采纳具体的数学技术方法，才能使教育评价建立在科学的基础上，作为教育工作者特别是从事教育评价工作的有关人员来说，了解和掌握教育评价中的技术方法是非常必要的。

本书对教育评价中的技术方法进行了系统的整理和归纳。全

书共分九章。从知识结构角度来看，大致可以作如下划分：第一章到第三章介绍描述统计和推断统计的基本方法，是全书的知识基础；第四章到第五章介绍教育评价中常用的重要数学方法，包括层次分析法和模糊综合评判法等，是全书的核心内容；第六章到第七章介绍如何对教育评价进行质量检查和结果分析，可以看做是对前面知识的综合应用；第八章介绍多元分析法，包括主成分分析、因子分析、聚类分析和判别分析，是对前面知识的某种深化提高；第九章对计算机统计软件 SPSS 的应用进行了详细介绍，既是对全书的一种总结，也为各种技术方法的应用指出了途径。最后是一些常用的统计表。

就编者所知，迄今为止对教育评价中的技术方法的阐述，尚散见于各种教育统计与测量、灰色理论、系统理论等方面的书籍文献中。编者在教学过程中深感不便，有意将它们梳理归纳，有机整合。在编写过程中特别考虑了适应一般读者阅读，在数学处理上去繁就简，尽量以例举证，良苦用心敬请读者明察。

教育评价的科学化是一个长期的过程，编者愿尽微薄之力，但限于学识水平再加之时间仓促，书中谬误在所难免，敬请读者不吝指教。

本书在编写过程中，得到了总参军训和兵种部董彦省参谋以及王从双、吉祥、赵昕、张婧、杨可为、许俊宏、庞博等研究生的大力协助，在此一并感谢。

作 者

2006 年 9 月

目　录

第一章　描述统计

单纯地对一组数据进行整理、概括，显现其面貌特征的统计方法称为描述统计。通过教育调查和教育实验获得大量数据，用归组，编表，绘图等统计方法对其进行归纳、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。描述统计的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。

第一节　数据的整理

研究教育问题时，要对研究对象进行调查、观察和测量，我们把搜集记录下来的数量依据称为数据，它以数字形式表现出来，如学生的分数、身高、体重等。

收集的原始数据，在未经整理时多是分散和零乱的。为了揭示和发现一组数据所反映的事物内在规律，需对这些数据进行科学的分类和整理。本节我们先简单介绍数据的特性、分类和抽样方法，然后介绍几种整理数据的方法。

一、数据的特性和分类

1．数据的特性

在教育统计中遇到的数据，首先具有离散性，如全年级的考试成绩不会全都相同，而是分散在一定范围内；其次是波动性（变异性），比如，一个学生的真实成绩是一定的，但几次测验的成绩不相同，总是在真实的成绩上下波动；再次是数据的规律性，教育统计中的数据经整理后呈现出某种规律，如一个班的学生考试分数或身高数据出现的频率，表现出中间大、两头小的规律。

2. 数据的类型

（1）离散型数据

一般指取整数值的数量指标，是计数性的（亦称计数数据），数据之间不能划分为更小的单位。如学校的个数、学生人数以及学生能力等级的量化，用5代表优、4代表良、3代表中、2代表差等，都属于离散型数据，只能取数轴上有限个不连续的值。

（2）连续性数据

一般指经过量度和测量而得到的数量指标，又称测量数据。如身高、体重、考试分数等。这类数据取值可以连续变化，数据之间能无限细分，通常可以用小数表示，可能取数轴上某个区间上的一切值。

二、样本及其抽样方法

1. 总体和样本

在统计学中把研究对象的全体称为总体（或母体），组成全体的每一个基本单位叫个体，而从总体中抽取的一部分个体称为样本（或子样），样本中所含个体的数目称为样本容量，通常用 n 表示，$n \geq 30$ 称为大样本，$n<30$ 为小样本，如全区初二5000名学生的数学成绩是总体，每个学生的成绩是个体，抽取100人的成绩为样本进行研究，则 $n=100$。

采用抽样调查获取数据时，随着样本选取的不同，结果也可能不同。因此，抽样时要求样本应能较好地代表总体，使样本成为总体的缩影，这样才能从样本的性质来估计和推断总体的相应性质。一般来说，样本容量越大，抽样误差越小，样本性质越接近总体性质。只要条件许可，比如抽样和计算都不困难，则样本容量越大越好。关于样本的大小在第二章还要论述。

在教育评价中大量采用抽样的方法，比如考察“教学质量”指标，就常常从一所院校中抽取一定数量的学生进行成绩测试。

2. 抽样方法

（1）随机抽样法

从总体中抽取样本时，每一个体都可能被抽到，而且有同等可能和相互独立的被选机会，这称为随机抽样。随机抽样的方法主要有抽签法和随机数表法。

抽签法是将所有个体编号，打乱次序后用类似于抽签的方法从中抽取样本。当个体数目较大时，抽签是不方便的，可用下面的查表方法。

利用查随机数表的方法取样，对 n 较大的情况是很方便的。随机数表是根据数理统计原理，由许多随机数字排列起来的数字表。表中数字的构造方法是：利用计算机使十个数字 0，1，…，9 中每次自动出现一个，它是不受主观意念支配而随机出现的，且出现的机会均等。这样得到一串数字，然后每五个数编排为一组，就可得到一个随机数码表（见附表 1）。应用时，将总体中的每一个体按顺序编号，然后用表中的不同数字来识别总体中的不同个体。

例如，我们想从 300 名学生的物理成绩中，抽取 n=30 的样本进行统计研究。首先将 300 名学生按顺序编号：001，002，…，300。然后，在随机数码表中取任意数字作起点，如随意投一枚针，把针尖所指的数字作为起点，假如针尖所指的是 04 行、01 列的数字 12。则从这个数字开始往右边取，到达右端后移至下一行继续往左取。由于学生 300 名是三位数，故取得三位数的顺序如下：

$\underline{125}$ 685 992 696 966 827 310 503 729
315 571 $\underline{210}$ $\underline{142}$ $\underline{188}$ $\underline{264}$ 981 765 559
563 564 385 482 462 $\underline{231}$ 624 …

凡在 000 至 300 之间的数入选；上列数字下边画横标者即为入选数字。大于 300 或重复的数舍去，直到取够 30 个为止，学生编号与入选数字相同者即为抽取的样本。

（2）等距抽样法

把总体中的个体按一定次序（如按座号或学号）排列后，按相等距离抽取样本。例如某年级有 180 名学生，我们要抽查部分学生的作业，先按座位编号，然后每隔 6 人抽 1 人。第一个人可从 1 ~ 6 号中随意取一个，如取 4 号，则接下去有 4，10，16，22…。

（3）分层抽样法

按某种特性，把总体分成若干层（或组），使用随机取样方法，选出与每一小组成比例的样本，作为研究对象。对于一所大学，学生的年级、专业不同，采用分层抽样更能反映整体情况，因而分层抽样是教育评价采取的常用抽样方法。例如某系四个专业共 200 名学生，抽取 20 人为样本进行研究，可以按各班学生学号，随机取 5 个学号，如 06，16，26，36，46，然后把四个班中具有以上 5 个学号的学生组成 $n=4\times5=20$ 人的样本。

再如，总体由 1355 份试卷组成，今拟从中用分层法按比例抽取 $n=100$ 的样本，其步骤如下：

①按分数段分为 10 层：0 ~ 9，10 ~ 19，20 ~ 29，…，90 ~ 100。

②统计各分数段（或层）个体占总体的百分比，设为：

0.67%，1.47%，2.73%，6.86%，10.33%，14.10%，17.41%，19.0%，18.52%，8.86%。

③按样本容量 $n=100$，依上列比例，确定在 10 个分数段中抽取的卷数分别是：1，1，3，7，10，14，17，19，19，9。合计为 100 卷。

④按以上确定的卷数，在各层中使用随机抽样法或等距抽样法抽取。如在 90 ~ 100 这一层中共有 90 卷，应取 9 卷，可依一定顺序每隔 10 卷抽一卷。

三、数据的列表法

对收集到的大量原始数据进行整理的方法之一是列表。列表就是对数据分别归类，用表格的形式表示出来，使之系统、直观，

并便于比较。一般有单项表、双项表、多项表和频数分布表等。

在一组数据中，每个数据出现的次数称为该数据的频数，按频数分类列出的表称为频数分布表，即对一组数据按其大小划分出等距的分组区间，然后将数据按其大小列入各相应的组别内。这种表在教育统计中是常用的表，下面我们具体介绍它的编制方法。

1. 简单频数分布表

现在通过具体例子说明编制的方法。某班 53 人，某次测验的分数如下：

69	56	78	89	68	99	83	83	61	76	62
69	63	68	83	57	66	67	45	89	78	79
89	94	74	79	59	73	88	52	72	87	94
65	90	70	87	69	89	67	93	53	74	54
70	59	64	73	72	60	75	75	65		

以上 53 个数据是杂乱无章的，可以通过下述步骤进行列表整理。

①求全距。原始数据中最大值与最小值之差即为全距 R。本例中 $R=99-45=54$。

②定组数。根据情况，一般定 10～20 组为宜。

③定组距。以组数除全距即组距 i，一般取整数。本例中若取组数为 11，则 $i=R/$ 组数 $=54/11=4.9\approx5$。

④定组限。组限是一个组在数尺上的起点和终点值。如：45～49，50～54，55～59，…，95～99。或 44.5～49.5，49.5～54.5，54.5～59.5，…。

⑤定组中值。组中值 x_c 是居于该组分数尺度上中点位置的数值。用组的实下限加以组距的一半即可求得。如 45～49 组的中值 $x_c=44.5+5/2=47$。本例中各组中值如表 1－1 中第 2 列所示。

⑥归类划组登记。用记数符号“正”或“+”把数据登记在适当组内，每组频数为 f，总频数为 N。本例的分数分布表即可制成，如表 1－1 中第 1、2、3 列所示。

2. 相对与累积频数分布表

（1）相对频数分布表

即由各组频数 f 对总频数 N 的比值所列的表。相对频数 $Rf = f/N$，其总和等于 1。如表 1－1 中第 4 列所示。

（2）累积频数分布表

可分为“以下”和“以上”两种累加频数的方法。在表 1－1 中把上限以下各组频数累加起来，如 49.5 以下为 1 人，54.5 以下 4 人，59. 5 以下 9 人，……。见表 1－1 中第 5 列所示，用 cf 表示。

表 1–1　53 人某次测验分数分布表

①组别	② x_c	③ f	④ Rf	⑤ cf	⑥ cf/N
95 ~ 99	97	1	0.091	53	1.0
90 ~ 94	92	4	0.075	52	0.98
85 ~ 89	87	6	0.113	48	0.90
80 ~ 84	82	4	0.075	42	0.79
75 ~ 79	77	6	0.113	38	0.71
70 ~ 74	72	8	0.150	32	0.60
65 ~ 69	67	10	0.188	24	0.45
60 ~ 64	62	5	0.094	14	0.26
55 ~ 59	57	5	0.094	9	0.17
50 ~ 54	52	3	0.056	4	0.075
45 ~ 49	47	1	0.019	1	0.019
Σ		N=53	1.00		

（3）相对累积频数分布表

用累积频数除以总频数，即 cf/N，见表 1－1 中第 6 列。若以相对累积频数乘以 100，就得累积百分数：

$$cp = \frac{cf}{N} \times 100$$

四、频数分布图

频数分布图是利用频数表中的数据找点、描线而做成的图可以从中清楚地看到数据间的关系和总体变化规律，直观形象，便于比较研究。

1. 频数直方图

在横坐标上标出等距分组点，纵坐际上标出频数，以组距为底边，以频数为高画矩形，矩形的面积表示频数分布图形。图 1－1 是利用表 1－1 中的数据绘制的频数直方图。

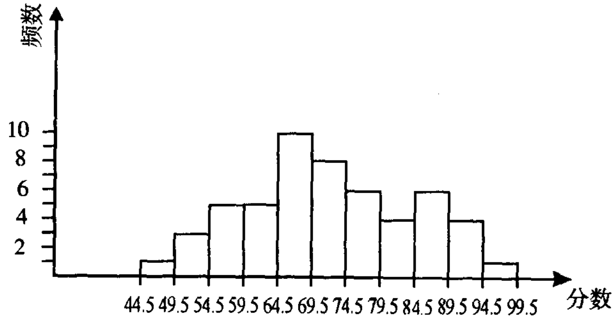

图 1–1　频数直方图

2. 频数多边形图

仍以横轴表示分数，纵轴表示频数，以每一分组区间的组中值为横坐标，以各组的频数为纵坐标点，连接各点成一条折线。为使折线的首尾都在横轴上，在原分组两端各加一个频数为零的组。图 1－2 是由表 1－1 绘制的分数分布的频数多边图。如果样本加大，组距缩小，就能绘出一条光滑曲线，称为频数曲线（修匀曲线）。

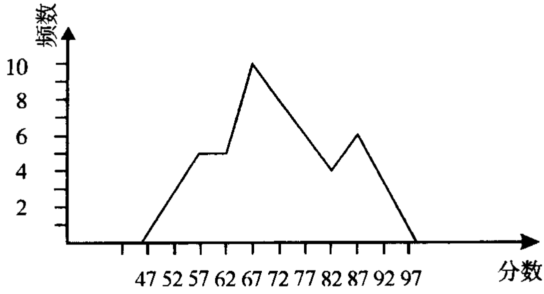

图 1–2　频数多边形图

3. 累积频数分布图

该图是横轴取每组上限，纵轴取累积频数，在相交处画点，顺次连接各点而成一上升曲线。如果以累积百分比为纵轴上点，重复上述过程，则得累积百分比分布曲线图。图 1 – 3 绘出了表 1 – 1 数据的累积百分比分布曲线图。这种曲线有一定实用价值。从图中可以看出 59. 5 分以下的人数占 17%，89. 5 分以下的人占 90%。若从纵轴 50 处画横线，与曲线交于一点，从这一点做横轴的垂直线，与横轴的交点约为 72 分。这说明 50%的人在 72 分以下，50%的人在 72 分以上。因此该图能较为准确地说明某个分数在全体分数中所处的位置。

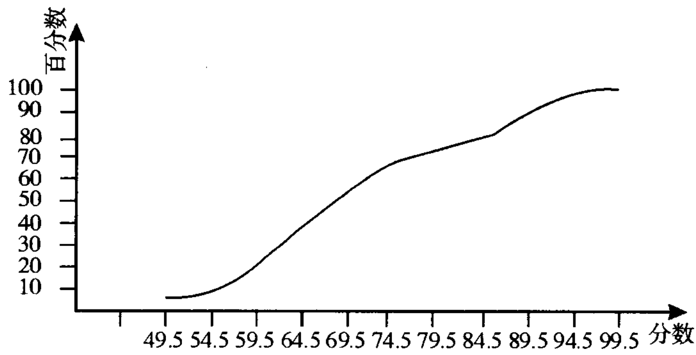

图 1–3　累积频数分布图

第二节 集中量数

频数分布表和频数分布图仅仅对一组数据进行了归纳性描述，但要对数据作进一步的分析研究，还需要计算出描述这一组数据特征的某些量数。例如，一组数据向何处集中?出现最多的数值是什么?其中间数值在哪里?这些能够反映一组数据集中趋势或一般水平的数值，统计学上称为集中量数或水平值。常见的集中量数有平均数、众数、中数。

一、平均数

平均数表示一组数据集中的位置，又称为均值。

1. 算术平均数

算术平均数是所有数据之和除以数据个数的商，记为$\bar{x}$。读为"x杠"。

（1）不分组数据求算术平均数

$$\bar{x}=\frac{1}{x}\sum_{i=1}^{N}x_i \qquad (2-1)$$

式中：x_i为第$i=1$， 2，…，N，N 为数据总个数

【例 1-1】 某校射击队 5 名队员在一次射击中，射中的环数分别为 6，7，8，9，10，求平均射中环数。

解 由（2－1）式

$$\bar{x}=\frac{1}{5}(6+7+8+9+10)=8 \text{（环）}$$

如果数据中有重复数，我们采用加权形式求算术平均数。"权"为所占的比重，比率，频率都可以看做为一种"权"。

例如，某校射击队 5 名队员在一次射击中射中的环数分别为 6，6，8，10，10，则

$$\bar{x}=\frac{1}{5}(6\times2+8\times1+10\times2)=8$$

把上式一般化得到

$$\bar{x}=\frac{1}{x}\sum_{i=1}^{N}x_i \cdot \frac{f_i}{N} \tag{2-2}$$

式中：f_i 为第 i 个数的频数，f_i/N 为第 i 个数的频率

我们称由（2－2）式定义的 $\bar{x}$ 为以频率为权的加权平均数，显然，权均为 $1/N$ 的加权平均数为算术平均数。

（2）分组数据求算术平均数（组中值法）

对于分组数据先要列出频数分布表，再把每组的各个数据都看做与组中值相同的数，这是因为每组内各个数据虽然有大有小，但其相对于组中值的误差最终趋于抵消，故可以把每组的组中值做为每组的代表值，由此得到

$$\bar{x}=\frac{1}{N}\sum_{i=1}^{r}m_i f_i \text{ 简记为 } \bar{x}=\frac{1}{N}\sum mf \tag{2-3}$$

式中：m_i 为第 i 组的组中值 $i = 1, \cdots, r$

r 为组数

f_i 为第 i 组的组频数

我们称由（2－3）式求平均数的方法为组中值法，由于我们假定每组中数据都与每组组中值相同，因此所得平均数结果不可能与将所有数据相加再除以数据总个数所得结果相同，利用组中值法求出的平均数只是一个近似值。

（3）$\bar{x}$ 的基本性质

常数性 $\overline{C}=C$，　　C 为常数；

齐次性 $\frac{1}{N}\sum cx=c\frac{1}{N}\sum x$；

可加性 $\frac{1}{N}\sum(x+y)=\frac{1}{N}\sum x+\frac{1}{N}\sum y$；

特别 $\frac{1}{N}\sum(x+c)=\frac{1}{N}\sum x+c$。