

教育方案评价丛书

如何进行统计分析

[美] 卡罗尔·T·菲茨-吉本 著
林恩·L·莫里斯
赵永年 纪明泽 江伯声 译
王 刚 校

上海翻译出版公司

Carol · T · Fitz-Gibbon Lynn · L · Morris

HOW TO CALCULATE STATISTICS

Sage Publications, Inc. 1978

本书根据塞奇出版公司 1978 年版译出

教育方案评价丛书

如何进行统计分析

(美) 卡罗尔 · T · 菲茨-吉本 著
林恩 · L · 莫里斯

赵永年 纪明泽 江柏声 译

王 刚 校

上海翻译出版公司

(复兴中路597号)

新华书店 上海发行所发行

沪江电脑科技排版公司排版 上海竞成印刷厂印刷

开本850×1156 1/32 印张6.5 字数173000

1989年4月第1版 1989年4月第1次印刷

印数1—4,000

ISBN7-80514-212-2

G.120 定价: 4.65元

中 译 本 序 言

这套《教育方案评价丛书》系根据美国加州大学洛杉矶分校评价研究中心主编的评价丛书译成。原文分为八册，译著亦同。各分册书名是：

- 《评价人员手册》
- 《如何处理评价目标》
- 《如何设计方案评价》
- 《如何测量方案实施》
- 《如何测量成绩》
- 《如何测量态度》
- 《如何进行统计分析》
- 《如何写评价报告》

丛书基本体现了评价工作过程的全貌，同时也大致反映了目前常见的评价方法的轮廓。

半个世纪以来，人们在世界各地对社会上从生产到研究的各种活动方案，广泛地进行了评价。用于评价的方法和技术也因此取得了较显著的进步。这种进展主要是从对教育方案——除了学校教育以外，还包括各种专业性的培训活动——积极进行评价的工作经验中产生的。在这段时期内，评价对象种类很多，但以各种教学改革方案和战时人员培训方案的目标及其实现程度为主要组成部分。

在这段发展过程中产生了一些重要经验，其中一条是：应当努力做到使评价有助于取得更多、更好的具有社会价值的效果，即通过评价提高活动方案的社会功效。——一句话，评价是为了提高。——这条经验主要来自教育评价。教育，作为嬗传人类文明并促使其加速进步的精神生产，必须经常进行适应于物质文明

和精神文明发展需要的改革，才能充分发挥提高人民素质，多出合格人才的社会功能。

在贯彻中共中央关于教育体制改革的决定过程中，全国正在积极开展教育评价工作。如：工科、师范高等院校等正在分别筹划对专业教学方案和学校管理水平进行评价的工作。从更广范围来说，全国正处在举世瞩目的改革过程中。改革的具体建议来自对过程和现况的评价。要做好评价工作，就必然会愈益讲究评价方法。

半个世纪来，世界各地在评价工作发展过程中积累了不少关于适应评价对象选用恰当的观察、分析和判断方式的经验，也就是选用评价模型的经验。他山之石，可以攻玉。国外经验中有不少可供参考借鉴之处。不过，当务之急，首先在于早让从事评价工作的广大干部迅速掌握基本方法。就这点来说，这套丛书的实效会是很显著的，值得向大家推荐。

这套丛书的译者多数是青壮年研究工作者，这是十分可喜的事。希望这些出版物的出现能引起更多人对于评价这门学科的重视，来共同推动有关工作和研究的开展，从而有益于我国社会主义改革与建设的进程。

邱 渊

1986年7月

目 录

中译本序言	1
第一章 统计导论	1
统计的用途	1
本书内容和格式	3
本书用途	5
进行统计分析	5
图表和概括统计量	6
差异的统计检验	7
统计的显著性检验	7
置信限	8
检定关系的统计: 相关关系	8
一些专用名词	9
第二章 对一组测量分数的概括	11
工作表格 2A 和 2B 的介绍	12
分数分布图形	12
什么时候作分数图	13
图形解释	15
工作表格 2A	20
工作表格 2B	23
工作表格 2C 和 2D 的介绍	26
计算平均数 $\bar{x}$	26
什么时候用平均数	27
工作表格 2C	28
工作表格 2D	30
工作表格 2E 的介绍	33

	计算标准差 S 和方差 S^2	33
	什么时候用标准差	34
	工作表格 2E	36
第三章	同一测量的两组分数间的差异检验	41
	统计的显著性检验用处何在	41
	如何在本章中找到适当步骤检验两组分数间的差异	48
	记住图形的结果	50
第一节	非配对组的分数比较	52
	工作表格 3A 的介绍	52
	两个平均数差异的统计显著性检验: 非配对组的 t 检验	52
	什么时候用 t 检验	54
	工作表格 3A	56
	工作表格 3B 的介绍	63
	对非配对组两个平均数差异置信限的计算	63
	什么时候用置信限	64
	工作表格 3B	65
	工作表格 3C、3D 和 3E 的介绍	72
	如果测量结果是秩次而不是分数的两组比较:	
	曼-惠特尼 U 检验	72
	什么时候用曼-惠特尼 U 检验	72
	工作表格 3C	74
	工作表格 3D	81
	工作表格 3E	87
第二节	配对组分数的比较	93
	工作表格 3F 的介绍	93
	两个平均数差异的统计显著性检验: 配对组的 t 检验	93
	什么时候对配对组用 t 检验	94
	工作表格 3F	95
	工作表格 3G 的介绍, 对配对组两个平均数的差异计算	
	置信限	103

	什么时候用置信限	103
	工作表格 3G	105
	工作表格 3H 的介绍	113
	如果测量结果是比较而不是分数的话: 符号检验	113
	什么时候用符号检验	113
	工作表格 3H	115
第四章	考察同组两次测量的关系	121
	相关系数 r 的含意	121
	什么时候不用 r	123
	如何选择合适的关系数	124
	关系数的解释	128
	r 的置信限	128
	以前可能已听到过的一句告诫语	131
	工作表格 4A 作图来描绘关系	133
	工作表格 4B 皮尔逊积差相关系数	137
	工作表格 4C 斯皮尔曼等级相关系数	141
	工作表格 4D ϕ 系数	147
	工作表格 4E 等级双列相关系数	152
	工作表格 4F 点双列相关系数	156
	工作表格 4G 决定 r 的置信区间	160
	工作表格 4H 的介绍	168
	χ^2 检验	168
	什么时候用 χ^2 检验	170
	在用 χ^2 检验时要避免的错误	172
	工作表格 4H χ^2 检验	174
第五章	用计算机分析数据: SPSS 系统	183
	用计算机进行数据分析应该了解什么	183
	帐户: 输出箱和口令的获得	184
	键控穿孔	185
	将数据输入计算机	186

在卡片上记录数据	186
准备一选卡片由计算机读数	189
为计算机分析准备数据	194
一些能用 SPSS 进行的统计分析	195
方差分析	195
偏相关	195
协方差分析	196
多重回归	196
判别分析	196
因素分析	198

第一章

统计导论

统计的用途

粗看一下词典，就会发现统计是一个试图使各种事实或数据集有序化的数学范畴。这是对统计能有助于从一大堆资料中得到有用数据的另一种解释。开展和经常运用统计的一个主要原因是，需要应付人类有限的操作记忆容量。任何人一下子面对着大量资料，都会很快茫然不知所措。因此，数据的概括和简化总是必要的。

进行统计的另一个目的是求出一些数据，这些数据能最精确地表达需要被描述的态度、成就和事件的性质。统计的目的在于精简观点、特性，并从一目了然的概括数字中对它们作出比较。然后讨论、记录这些数据，并以此作为作出决策，形成观点，发展理论的基础。

当然，运用统计需要知道它们的含义、认识到它们的局限性，本书旨在指导读者运用统计方法概括收集和报告的数据，对统计的目的和解释提供一些简要的说明。

统计在教育中主要具有三种意图。大部分评价都提供了实现这三种意图的机会。

1. 统计概括资料 概括统计量有两种，即集中量和差异量。平均数，中数和众数称为集中量，它有助于用一个数字描述一组数字分数的集中程度。而诸如标准差之类的差异量，则是通过反映对集中量起作用的分数是如何变化的来描述集中

量。差异量越小，平均数或中数的代表性就越大。最好把标准差看成是描述分数围绕平均数变化的平均量度，大的标准差表示与之相对应的平均数不能被看成是全组分数的一个好的概括量。

2. 统计告诉人们如何认真地看待方案或处理结果之间出现的各种外观的差异。如果准备报道某一事件，比如成绩分数或对态度问题的回答发生了变化，就要确认所发现的变化是真实的，而并不能归因于分数自然产生的波动。统计将帮助人们决定研究或方案评价所产生的结果是否可能由偶然因素所致。统计检验有一个习惯的做法：首先假定正在检查的方案或处理实际上并没产生任何效果。在这种情况下，如果结果显示了高度异常，便可证实无效果的假设必然是错误的。举例来说，从作无效果的操作假设开始，如果研究产生的结果有 75% 的时间是由于机遇产生的，那么无论是评价者还是评价听取人都不会认为该结果会影响所作出的决定。另一方面，在无效果的假设前提下，如果所得的结果由于机遇而产生的可能性很小，也许只有 5% 的可能，那么该结果将是可信的。
3. 统计帮助人们确定数据组间的关系程度。关系程度通常由一个相关量来体现。相关是表示两组分数一起变化的程度。举例来说，也许需要了解学生的年末数学成就分数是否与问卷表所表达出来的学生对学校的态度有关，通过计算这两组数据的相关，就能得出下列三条结论中的一条。
 - (1) 两组数据之间存在高度正相关。这意味着通过统计计算，所得的相关系数是个小于 1 的较大小数，说明随着一种测量分数，如数学成就测验的增大，另一种测量分数也相应增大。
 - (2) 两组数据间不存在相关。这意味着相关系数是一个和零无显著差别的小数。

(3)两组数据之间存在高度负相关，即相关系数是一个接近于-1的小数，这说明随着数学成就测验分数的增大，对学校的态度反而变得更差。

本书内容和格式

展现集中量、统计检验和相关是统计学的主要目的，也是本书统计内容的唯一重点。作者已尽力使本书只包括最基本、最常用的（统计）技术。本书所概述的方法并不仅仅适用于答复评价时所提出的一些最基本的问题，而且提供大体理解统计知识的坚实基础。如果谁有机会在生活经历中作为一个评价者，在特殊的情况下，也许需要使自己了解更复杂，更现代的统计。

本书内容以洛杉矶加利福尼亚大学评价研究中心的评价者的经验、教育测量和统计专家的建议以及在实际中使用过统计方法的学校教学人员的意见为基础。尽管计算统计量的公式是固定的，本书还是努力试图运用最直观的可理解的计算方法，并尽可能规定出程序以便理解统计程序怎样处理信息去产生一个特定的结果。

《如何进行统计分析》是一本工作手册。该手册的三个主要部分（第二、三、四章）是围绕着工作表格而组织起来的。每一个工作表格提供了计算和解释一个特定统计量的步进指导。每一工作表之前是一个《工作表格介绍》，解释统计什么和什么时候运用这一统计。

本书共分为五章。第二章是概括一组分数的程序，即集中量和差异量求法。在有一组分数变量需要分析和报告的任何时候，便常常可运用这些程序。第三章描述了在两组结果之间作比较的统计量，诸如实验组与控制组之间的成就测验平均分数。第三章中的统计量通常涉及到显著性检验，因为它们有助于确定如何认真看待出现在组间的成就、能力和态度的外观差异。除了进行统计检验的工作表格外，第三章还包括了称作“统计的显著性检验

用处何在”的一节内容。

第四章涉及到关系和相关。当需要知道某组一次测量所获得的分数与同组的另一次测量所获得的分数是否具有规律性的关系时，要用到这部分内容。举例来说，人们也许要了解孩子们的数学分数是否与这些孩子所接受的特定数学方案的每周训练次数有关。

第五章介绍了如何利用被最广泛应用的社会科学现成程序统计库（SPSS）作为实例，为计算机分析准备数据。这一章也描述了 SPSS 中的程序内存与本书中所介绍的统计量之间的关系。同时，第五章还包括对其他统计程序的简单描述，如果能利用计算机的话，这些程序可用来进行分析。

当然，如何运用第二章至第四章中的工作表格，应因人而异。大部分表格在书中已得到运用。然而，有些表格可能应当称为“指导表格”，因为评价者将不得不把表格或图表复制到自己的纸上，而不是应用工作表格本身。因为篇幅有限，本书无法列出大量的数据。

请注意，工作表格分为两个部分，步进指导表格和与之相对应的完成的实例表。还须注意，所举的这些例子是虚构的，是为了说明如何运用这些表格而提出的。由于这一点，所以例子经常是代表了一种无障碍情境中的理想化数据。如果实际所得的结果看起来不同于例子中评价者所获得的虽幸运然而却是虚构的结果，请不要失望。重要的是根据数字所得到的结果应讲得通，而这些数字则可以用书中提供的程序计算出来。

如果原来已选择了合适的统计方法，而且重复验算了计算过程，那么最初看来可笑的结果也许是正确的。当然，如果还有疑问，就让数据分析员来检查这一工作。

在应用第二、三、四章中的工作表格时，是以假定人们具有一个很简便的，带有开方键的电子计算器为基础的。为了有助于那些几年前就结束数学学习的人们，工作表格中只要求具备最初步的数学知识。由于人们对把数字代入公式常常感到困难，因

此，把计算分解成一些子步骤。当然根据每个人的数学程度，可跳过一些不必要的子步骤。

顺便说一下，第五章的内容是假定读者过去从未使用过计算机。

本书用途

因为读者需要或希望了解更多的统计知识，才会阅读本书。如果为了提高专业而独自阅读本书，应该常常有思考地仔细阅读本书的几小段，而不要试图一下子就领会全书。当读到各种不同的分析方法时，努力想象几种自己本职工作范围中能够应用这些方法的情境。不管获得怎样的数据，都尽可能运用这些程序。考虑一些具体的情境，这将使读者进一步熟悉本书中论述的程序，从而不久就能够很快选择与运用这些统计方法。进一步说，掌握本书中的这些基本技术也可为学习更复杂的统计方法作好准备。

如果探讨统计是为了马上能运用，那么首要任务是组织数据为统计分析作准备，并且确定这些数据正是所要找的和处理的。阅读下节，将对读者有所帮助。

进行统计分析

假定评价或实验研究已经到了这一地步；评价工具已经被设计（制作或购买）和实施。评价工具所完成的记录一经返回来，就开始数据分析。如果要由计算机完成一部分或全部分析工作，那么必须立即阅读第五章。如果用手工进行分析，那么，第二章几乎一直可作为最初的向导。需要对结果作出一个容易理解的精确的概括。一般来说，这种概括采取下列两种形式：

1. 一个或一个以上的图表
2. 一个或一个以上的概括统计量

图表和概括统计量

为描绘图形，先要准备一张数据概括表。这是一张记录着评价工具所得分数的纸，这样，在进行计算时，便能容易得到这些数据。如果所有分析均由计算机完成；为了便于键盘操作，数据须用标准编码纸进行概括，或者由仪器直接在纸带或卡片上穿孔。

一旦分数以一定的方式组织起来，就可利用工作表格 2A 或 2B 作出粗略的示意图，这种草图的目的并不是为报告提供一个概括量，而是提供了一个了解数据全貌的机会——寻找用什么方法进行分析的趋势。

计算概括统计量，也要求事先准备一个数据概括表。然后再选择是计算平均数，还是找出中位数。第二章对中位数有详细的描述，可以很快从图形或顺序排列的成绩表中找到。顺序表是一组从大到小排列起来的分数序列。除中位数外，还能指出较高和较低的四分位差（见第二章），以显示分数的展形或差异程度。这给读者一个中位数的典型的概念。

工作表 2C、2D 对计算平均数花了较长的篇幅。但是，通常人们还是优先选择平均数，因为在进行统计检验时常常需要。工作表格 2E 提到了标准差，它伴随着平均数，揭示了分数的展形。

计算了概括统计量以后，还应做些什么就由数据收集的目的所决定：

1. 如果研究所得出的结论取决于已发现的不同组的态度或成就的差异，那么就能运用统计数字向评价听取人至少提供一部分判断所存在差异的重要性的依据。通常这意味着进行统计的显著性检验。不过差异的统计检验的另一种方法——计算置信限，也是可取的，同样值得介绍。
2. 如果希望检查学生或方案的各种不同特征是否互相有关，那么，将运用相关方法估算一下这种关系的强弱。

差异的统计检验

如果已经采用一个设计，把组分为实验组和控制组来加以实施，那么在确定实验方案或程序获得什么利益时，研究已有一个内在的依据：控制组表明不实施方案会得到什么结果。通过检验，实验组和控制组的后测结果的偏差就能回答这一关键问题：方案是否造成差异？统计的注意力集中于得到的差异上，说明它是否大到足以支持两组之间确实存在差异的结论。

可以同时利用统计与设计来仔细核对方案实施前后组内的差异。注意方案或实验开始阶段的态度或成绩差异，能有助于确立实验组和控制组的初始等值的情况，如果希望解释研究结论中出现的差异，这是关键的一环。

在缺乏设计和在一定限制条件下，同样能运用统计检验检查同组预测与后测分数之间的差异，或者检查两组（比如男生和女生）在同一测验中是否存在差异。

如果希望比较两组的同次测量或一组预测与后测的分数；那么，就要进行显著性检验或计算一组置信限。

在有些情况下，可能会发现需要比较两组以上的结果。举例来说，可能会是这种情况，要对两个实验方案和一个控制组作出比较评价。在必须比较两组以上的结果时，需要用方差分析，第五章对方差分析作了简要介绍。大部分方差分析对于手工计算来说，即使有计算器的帮助也是太复杂了。所以方差分析的工作表格没有列入本书。如果读者认为方差分析对自身的研究工作适用的话，那么，应该找一个数据分析员用计算机来作分析。地方学区办公室或者一个地方大学的教育或心理学的研究生能很快进行方差分析。

统计的显著性检验

研究者常常以观察差异是不是一个重要的统计显著性检验上的差异，来作为判断的根据。虽然根据数据的不同特性只能为

数不多的不同检验，但是所有统计显著性检验都遵循一个共同的程序。

1. 结合统计理论和计算，用一张表格或图形来表明在两组没有真实差异情况下，所得各种范围的观察分数差异的概率。
2. 把这张表与所观察的差异作比较，判断该结果是否小得足以得出结论：所发现的可能确实代表了一个真实的差异。

第三章的导言部分，对统计检验的逻辑基础作了更详细的说明，同时还讨论了统计的显著性对评价者来说意味着什么。

置信限

虽然显著性检验表明了分数间的差异是否大到足以超出了随机所致，而置信限却为在重复进行研究时，实验组和控制组之间将得的差异提供了一个估计。这样，对使用者来说，置信限比显著性检验提供了更多的信息。它不仅表明可能发现的真实差异，而且还提供了有关差异大小的观念。

更进一步说，在计算置信限时仍然可以进行统计显著性检验。因为置信限用所得差异来描述两组分数间可能会发现的差异范围，一个统计显著的所得差异也就是一个置信限中排斥零差异的差异。这一推理在第三章有充分的阐述，指出了当要比较两组成绩或态度时，几乎在所有情况下，都可计算置信限并作统计显著性检验。如果有合适的的数据以及如果评价听取人所受教育已能达到了宁愿选择和运用置信限的程度，那么就应尽一切可能用。

检定关系的统计：相关关系

除了(或者并不)关心所作的研究是否使组的成绩产生差异之外，评价者也许想知道学生、教师或方案的特性之间是否存在关系。在这种情况下，感兴趣的是计算一种相关。相关是专门用来回答下列这类常见问题的：

1. 学生、教师或已实施过的两次测量的成绩之间是否有关系？
2. 已实施的测量和学生、教师或学校的一定特性（如他们的社会经济状况）之间关系的强度如何？

另外，还可有更特殊的理由计算一种相关。下列这些问题表达了与编制测量工具（如成就测验、态度调查和观察表等）有关人员所关心的问题：

1. 一个测验或一个态度工具的信度如何？
2. 某工具的预测效度或同时效度如何？
3. 两个观察者的报告可信度如何？

如果打算计算相关系数，请翻阅第四章。请注意，对相关也能计算置信限和显著性检验，以便决定对相关应相信到什么程度。

一些专用名词

统计工作要求精确，这不仅体现在数据的采集和计算方面，而且体现在用词方面。本书的程序中一再读到分数、工具、个案和组。这些术语中的每一个术语都是一个包罗万象的词，用来代表一整张表或短语，使讨论不会变得过于复杂。

分数和数据 数据最基本的组成是分数。一般说来，分数是代表学生一次测试中成绩的数字。但是，有时候要分析的数据不是由一般意义上的分数所组成。举例来说，数据可以由观察到的孩子参加工作次数的百分比组成；或者由教师所提供的等级，家长到会人数，或学生在班上的名次组成。但本书中只简单地用分数这个词，而不使用信息所取各种形式的名称，那么，请注意分数的广义解释，它意味着由测量工具产生的任何数字。

工具、测量工具或测量 这些术语是指获得信息的任何手段。下列这些都被认为是测量工具：购买的或当地制作的成就测验、观察表、面谈、记录、报告、检查表和问卷。

个案 和分数一样，是评价者常用的统计量，来自个人——