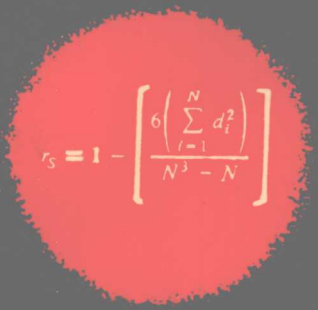

余嘉元

NONPARAMETRIC METHODS

心理和教育研究中的非参数方法

大连海事大学出版社

心理和教育研究中的非参数方法

余嘉元　编著

大连海事大学出版社

(辽)新登11号

内 容 提 要

本书介绍了40种非参数统计方法，它们计算简便，对数据分布没有严格的要求，可运用于小样本的情况，很适用于心理和教育研究。作者把这些非参数方法按适用于单个样本、两个样本和多个样本等不同情况进行分类介绍，每一种方法都配有实例，使读者很快就能进行实际的运算。

本书可供各级各类学校的教师、师范院校的学生以及心理和教育科学研究人员学习使用。

心理和教育研究中的非参数方法

余嘉元　编著

责任编辑：明凯　　封面设计：余乐孝

大连海事大学出版社出版、发行

南京师范大学印刷厂印刷

开本：787×1092　1/32　　印张：13.25　　字数：287千字

1994年8月第1版　　1994年8月第1次印刷

印数：0001～1000　　定价：9.80元

ISBN 7－5632－0591－8/G・89

目　录

第一章　绪　论

第一节　统计检验概述

一　心理和教育统计

无论是心理学工作者还是教育工作者，常常会遇到大量数据，如课堂测验、期中和期末测验、毕业考试、升学考试、能力测验、人格测验和问卷调查等，其结果总是令人生畏的复杂的大量数据。科学研究的目的决不是仅仅为了得到这些数据，而是希望通过数据了解事物的本质属性、内在规律和事物间的相互关系。在心理和教育研究中，由于被研究的对象是具有高度复杂性的人，而且，我们还无法非常精确地把人的各种心理量度量出来，这就使所得到的结果具有很大的随机性。如何把搜集到的数据进行整理、排列、分类、制成量表或找出这些数据的代表性量值，使我们能够用简洁的方式或量值表示大量纷乱的数据，这就是描述统计的任务之一。

当然，这还只是对于一组数据而言，当研究中有两组或多组数据，而我们又要了解各组数据间的相互关系时，就要计算它们之间的相关系数，这也属于描述统计。由于描述统计往往是我们处理数据的第一步，因此，它是教育和心理统计中最基础的内容。

但是，人们对于客观世界的认识，决不会满足于描述统计的结果，心理和教育工作者还希望通过对部分对象的研究而获得对全局的认识。显然，在心理和教育研究中，人们很难对

全体对象逐个进行分析，但又希望根据部分数据而得到全体的资料，或者要对两种事物或多种事物之间的差异进行比较，这就是推论统计的研究课题。心理和教育研究中的大多数课题都涉及到推论统计，因此人们把它称为心理和教育统计中的核心问题。

随着科学的发展，人们在认识客观世界的过程中，往往会提出各种假设。为了要验证这些假设，就需要用科学的方法来设计实验，这就是统计学中近几十年发展起来的实验设计，由此可见，描述统计、推论统计和实验设计是统计学的三个组成部分。

到目前为止，统计学中的大部分方法都是建立在对总体的性质作出了许多假设的基础上的，例如，总体是正态分布的，两个总体之间满足方差齐性。由于我们把代表总体的值称为参数，因此这些方法称为参数统计方法。然而，在心理和教育研究中，有许多总体的分布并不明确，或者不符合参数统计的前提，这些问题就不能用参数统计的方法进行处理。于是统计学家又发展了一类新的统计方法，该类方法的共同特点是对数据的总体分布不作严格的假设，也就是说，不论是何种分布形式的总体，都可以运用这些方法，这类方法就被称作非参数方法或"分布自由"的方法，它极大地扩展了统计学的应用范围，在心理和教育研究中有着广阔的应用前景。

二　非参数统计方法的特点

综上所述，非参数方法的最大特点是不需要考虑总体的分布形式，这就大大方便了心理和教育工作者。因为对于总体分布形式的考察是一件需要花费一定精力的工作，而且，目前的大多数统计方法只适合于总体为正态分布的情况，对于其

它分布形式的总体没有提供很多有效的统计分析工具，非参数方法则是完全打破了这一框架，总体分布形式对该方法并不起任何限制作用。

非参数方法的另一个特点是它并不要求数据具有等距性质，也就是说，它并不要求数据能进行四则运算，而只要它们具有分类或顺序的性质就可以了，即只要研究中的变量是名义变量或等级变量，都可以运用非参数方法。在心理和教育研究中，恰恰是名义变量和等级变量用得最多（要得到真正的等距变量并不是很容易的），因此，非参数方法是心理和教育工作者进行科学研究的有力工具。

非参数方法还有一个特点，就是其计算方法相当简单，不需要进行繁琐复杂的运算，不需要掌握高深的数学知识就能进行操作，这也是该方法受到普遍欢迎的原因之一。

非参数方法的另一个引人瞩目的特点是可应用于小样本的情况，这也为许多研究工作者提供了方便。在运用参数统计方法时，为了得到可靠的结果，一般需要从大样本获得数据。但是大样本的研究往往费时费力高成本，有些研究还可能很难得到符合要求的大样本，例如对于双生子的研究，就很难获得大样本，这就给有关的科研工作带来了困难，非参数方法对于小样本数据的有效性，为克服这一困难提供了帮助。

尽管非参数统计方法有如此众多的优点，但国内还没有一本由我国学者自己撰写的、供心理和教育工作者使用的关于非参数统计方法的专著，这实在是一大遗憾。虽然国内也有一定数量的非参数统计方法的著作，但它们多数是以专业数学工作者为读者对象，高深的数学推导使许多心理和教育工作者望而却步。有些心理和教育统计学教材中虽然介绍了一

些非参数方法，但还不够全面，因此出版一本实用性较强的、供心理和教育工作者使用的非参数统计专著，是很有现实意义的。本书并不涉及到繁复和高深的数学推导，使大多数心理和教育工作者都能看懂。另外，本书的内容是根据实际工作者在研究中遇到的具体情况而进行的，例如，根据样本的不同（单个样本和多个样本、相关样本和独立样本），变量的不同（名义变量、等级变量和等距变量），介绍不同的非参数统计方法，并且，对于每一种方法都阐明了它的原理、功能、计算方法和具体步骤，还给出了实际例子，因此，对于我国的心理和教育工作者将有一定的实用价值。根据西格耳等国际统计学权威的经验，对于每一个例题都运用了相同的六个步骤加以说明，这将有助于读者学习和运用这些非参数假设检验方法。

三　统计检验的步骤

在进行心理和教育统计检验时，通常按照以下六个步骤加以实施：

1　建立假设

在进行某项研究之前，心理和教育工作者往往会提出某种假设，然后进行实地调查、测量或实验，并且根据所得到的资料来检验该假设的真实性，这种假设称为科学假设。例如，在研究初中男女学生的数学能力以前，研究者可能会根据以往的经验和已有的资料，假设男生的数学能力高于女生，然后对男女学生进行相同的标准化数学能力测验，若所得资料支持了原先的假设，那么该假设就得到证实，否则该假设就应予以拒绝或舍弃。如果用统计的术语或符号将假设加以陈述，那么就把它称为备择假设（或研究假设），记作 H_1。例如，研究者认为“初中女生的数学能力低于初中男生的数学能力”，这是

科学假设。若运用统计的述语，则可以陈述为“初中女生的平均数学能力(μ_1)低于初中男生的平均数学能力(μ_2)”，这就是备择假设，用符号表示就是 $H_1:\mu_1 < \mu_2$。

作为研究者，本来应该搜集资料来支持或证明这一假设的真实性，但在统计学中却一直是沿用费歇尔的方法，即首先建立一个和备择假设相对立的假设，即虚无假设，记作 H_0，上述例子中的虚无假设为 $H_0:\mu_1 \geqslant \mu_2$，然后，运用反证法来检验备择假设，即研究者设法搜集支持或证实虚无假设的材料。当研究者所获得的证据不足以支持虚无假设时，则 H_0 就被拒绝，备择假设 H_1 就成立；反之，当研究者有足够的证据支持虚无假设时，则 H_1 就被拒绝。由于 H_1 是 H_0 的对立面，人们也把 H_1 称作对立假设。在统计学中，往往把 H_0 写在前面，把 H_1 写在后面。

作为统计检验的第一步，就是要建立虚无假设 H_0 和备择假设 H_1，同时还要决定应该在哪种条件下搜集哪些资料，以便进行统计检验，这就是要确定哪些是研究中的自变量、因变量和控制变量，各变量的数据是取自于独立样本还是相关样本，以及这些变量是属于何种性质，是名义变量、等级变量还是等距变量。

值得注意的问题是：当 H_0 被拒绝时，虽然我们可以说所搜集的资料支持了备择假设 H_1，从而接受这一假设，但并不能由此而认为 H_1 就肯定是正确的，因为统计推理只是从概率论的角度看待问题，在对结果进行解释时，我们只能认为 H_1 为真实的概率较大。

2　选择统计检验的方法

迄今为止，统计学家已经提出了许多种统计检验的方法，

每种方法只适合某些特定的条件，而且，都必须满足于一定的假设。在选择统计方法时，应该考虑到测量的水平（即数据的类型）、样本的数量（包括单个样本、两个样本和多个样本）、样本间的关系（相关样本和独立样本）等多方面的因素，我们将对该问题作详细讨论，本书的附录中列出了适用于各种条件的非参数统计方法，可以供读者参考。

3　显著性水平

统计检验的结果可能是接受虚无假设 H_0，也可能是拒绝 H_0，但无论是哪种结果，我们都无法完全确定 H_0 的真伪，而只能说明其真实程度的高低，或者说以某种概率支持或拒绝虚无假设 H_0。在根据统计数据作出结论时，无论何种推论都有可能犯错误，这些错误可以分为两类：第一类型错误（又称 α 错误）和第二类错误（又称 β 错误）。

当虚无假设 H_0 为真而遭到拒绝时，研究者所犯的错误称为第一类型错误，其犯错误的概率 α 称为显著性水平。在心理和教育统计中，α 值的确定带有一定的任意性，通常采用 $\alpha = 0.05$，$\alpha = 0.01$ 或 $\alpha = 0.001$ 作为显著性水平。当 $\alpha = 0.05$ 时，研究者作出结论时所犯第一类型错误的机会为百分之五，其余类推。α 值越小，犯第一类型错误的概率越小，即研究者确定的显著性水平越高；α 值越大，犯第一类型错误的概率越大，即研究者确定的显著性水平越低。

当虚无假设 H_0 为假而被接受时，研究者所犯的错误称为第二类型错误，其犯错误的概率记作 β，那么，$(1-\beta)$ 就是当 H_0 为假时，它被正确拒绝的概率，通常称为统计检验力。从理论上讲，为了对研究结果进行判断，必须事先确定 α 和 β 的值，但研究者通常习惯于先规定 α 和 N，由于它们和 β 之间

有某种确定的数学关系，因此当 α 和 N 确定后，β 也就确定了。当 N 给定时，α 的减小将导致 β 的增加，而 β 的减小又将使 α 增加，两种类型的错误不可能同时减小。如果既要降低 α 的值，又希望减小 β，唯一的方法就是增加 N，这也是许多统计方法要求取较大被试样本的原因。

4 *抽样分布*

在选定了统计检验方法、显著性水平和样本容量以后，就应该考虑该检验统计量的抽样分布，因为当取相同的显著性水平和样本容量时，对于不同的统计检验方法，由于其检验统计量的抽样分布不同，其临界值也是不同的。

抽样分布是样本统计量的理论分布。假设我们从初中学生中随机抽取了 3000 人的样本进行韦克斯勒智力测验，得到平均分数 $\bar{X}_1 = 105.8$；然后再从同一个总体中抽出另一个 3000 人的样本，进行相同的智力测验，得到平均分数 $\bar{X}_2 = 103.6$；假设连续抽了 1000 个都包含有 3000 人的样本，将这些样本的平均数作一个次数分布，这就是样本平均数的经验性抽样分布，该分布表示了当 H_0 成立时，样本平均数可能取各种值的概率。但实际上，任何一个研究者都不可能抽取两次、三次，甚至一千次样本，抽样分布的形式通常是从理论上推导出来的。任何一个统计量（如平均数、标准差、相关系数、百分比）都有其相应的抽样分布，而且其中多数是服从正态分布的，每一个抽样分布都有其标准差，称为该统计量的标准误。当一个变量为正态分布时，其分布形式可以由平均数和标准差完全确定（参见附表 1）。

5 *拒绝区*

拒绝区是抽样分布的一个区域，当 H_0 为真时，我们实际

观察到的样本统计量落入该区域的概率为 α。在进行统计假设检验时，究竟是接受还是拒绝虚无假设，通常要看所得到的样本统计量是否落入该区域。

由于研究者的统计假设 H_1 有时具有方向性，有时没有方向性，因此，拒绝区可能是单侧的，也可能是双侧的。例如某个心理学家认为初中男生的平均数学能力 μ_1 高于初中女生的平均数学能力 μ_2，那么其相反的陈述就是虚无假设 H_0，即 $H_0:\mu_1 \leqslant \mu_2$，而备择假设为 $H_1:\mu_1 > \mu_2$。从该假设中可以看出，研究者强调了男生的平均数学能力 μ_1 高于女生的平均数学能力 μ_2，也就是考虑了这种差异的方向性，这种具有方向性的统计检验称为单侧检验。在一般情况下，若 $H_1:\mu_1 < \mu_2$，表示研究者只强调第一个群体的平均数小于第二个群体的平均数，而对于 $\mu_1 > \mu_2$ 的情况不予考虑，这时，拒绝区在分布的左侧；若 $H_1:\mu_1 > \mu_2$，表示研究者只强调第一个群体的平均数大于第二个群体的平均数，而对于 $\mu_1 < \mu_2$ 的情况则不予考虑，这时，拒绝区在分布的右侧。

如果研究者在提出备择假设时，只考虑了两个群体平均数之间的不相等，而没有强调哪一个群体的平均数较大或较小，即没有强调差异的方向性，这时就必须在分布的左右侧同时加以考虑，这种不考虑方向性的检验称为双侧检验。例如，研究者只强调初中男生和女生的平均数学能力不相等，而不考虑他们之间哪个群体更高，因此，其假设为

$$H_0:\mu_1 = \mu_2$$

$$H_1:\mu_1 \neq \mu_2$$

这时，拒绝区就要分成相等的两部分，它们在分布的左侧和右侧各有一半，面积均为 $\alpha/2$，也就是说，当显著性水平 α 确

定时，双侧检验的拒绝区比单侧检验的拒绝区离平均数更远。例如，取 $\alpha = 0.05$，若采用单侧检验，其拒绝区离开 Z 分数平均值 0 的距离为 $1.645Z$（我们把它称为临界值），若采用双侧检验，临界值就是 $1.96Z$。

在某个实际研究中，究竟是运用单侧检验还是双侧检验。一定要根据研究目的来作出决定。通常，当备择假设中有“大于”、“小于”、“高于”、“低于”、“比 …… 好”、“比 …… 差”等涵义时，就采用单侧检验；当备择假设中有“相等”、“不等”、“同样”、“不同”等涵义时，就采用双侧检验。

6　判定

在完成了上述各步骤以后，就可以搜集资料，并根据从样本得到的数据，计算出所需的统计量，就能很方便地作出判定。如果所得到的统计检验量落在拒绝区内，此时虚无假设为真的概率很小，就可以据此作出判定：虚无假设 H_0 为假，应该接受备择假设 H_1，作出这种判定时，犯第一类错误的概率为 α。

下面举些例题[1]，主要用以说明怎样运用本节中介绍的统计检验的六个步骤来作出统计判定。

〔例题 1.1〕　假设有位研究者怀疑目前农村中男女儿童比例不平衡，男孩的人数可能超过女孩。为了检验这一假设，他到某地农村去随机抽取了 12 个儿童[2]，观察男孩出现的频率，并

---

[1] 为了让读者更好地掌握本书阐述的各种方法及实际使用步骤，作者给出了一些例题，其中有些例题参考或引用了 Siegel, S. & Castellan, Jr. N. J. 的 Nonparametric Statistics for Behavioral Sciences, (2nd ed. 1988) 一书中的有关题目，有些例题是作者自编的。

[2] 这里只是为了说明本节中介绍的统计检验的六个步骤，故样本较小。在解决类似的实际问题时，应取较大的样本。

按以下步骤进行统计检验：

〔1〕建立假设

$H_0: p(B) = p(G) = 0.5$

该假设表示，对于农村儿童而言，出现男孩的概率 $p(B)$ 和出现女孩的概率 $p(G)$ 之间没有差异，即男女比例是平衡的。

$H_1: p(B) > p(G)$

该假设表示，男孩的比例高于女孩的比例。

〔2〕统计检验

由于在该研究中，变量（儿童）只可能取男孩或女孩两种值，故其合适的检验方式为二项检验。

〔3〕显著性水平

研究者决定采用 $\alpha = 0.01$ 作为其显著性水平，根据题意 $N = 12$。

〔4〕抽样分布

抽样分布是指 H_0 成立时，得到 x 次男孩和 $N - x$ 次女孩的概率，其分布由二项分布函数给出，即

$$P(x) = \frac{N!}{x!(N-x)!}\left(\frac{1}{2}\right)^N,$$

$x = 0,1,2,\cdots,N$。

表 1.1 列出了此时出现男孩和女孩的抽样分布。

从该抽样分布可以看出，当 H_0 成立时，在随机抽取的 12 个儿童中，最可能得到的结果是 6 个男孩和 6 个女孩（其出现的频数为 924），出现 7 个男孩和 5 个女孩的频数稍小，但还是很可能的，而出现 12 个男孩或 0 个男孩的可能性就极小。