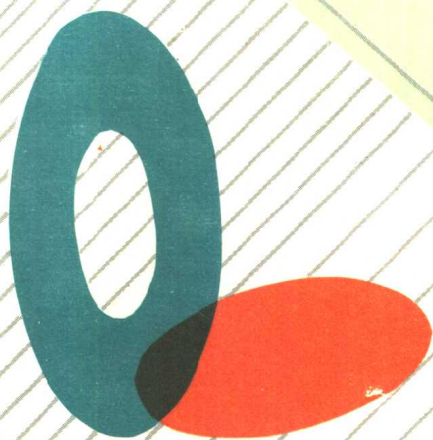


非参数统计

——方法与应用

易丹辉 编著



中国统计出版社

非参数统计

——方法与应用

易丹辉 编著

~~中国统计出版社~~

(京)新登字 041 号

图书在版编目 (CIP) 数据

非参数统计: 方法与应用/易丹辉编著.

—北京: 中国统计出版社, 1995. 6

ISBN 7-5037-1968-0

I. 非…

II. 易…

III. ①非参数统计—方法 ②非参数统计—应用

IV. O212.7

中国统计出版社出版

(北京三里河月坛南街 38 号 100826)

新华书店经销

中国科学院印刷厂

850×1168毫米 32开本 9.625印张 21.5万字

1996年3月第1版 1996年3月北京第1次印刷

印数: 3500册

定价: 14.00元

(版权所有 不得翻印)

前 言

非参数统计方法是本世纪 30 年代中后期开始形成并逐步发展起来的。它是与“参数统计”相比较而存在，不依赖于总体分布及其参数，亦即不受分布约束的统计方法。作者在承担国家教委人文社会科学研究八五规划项目博士点基金项目“90 年代我国居民消费结构及倾向的研究”过程中，感到在很多情况下，参数统计方法的运用受到限制，如研究居民消费行为、居民收入等级与消费结构的关系等等，从而转向了非参数统计方法的探讨。《非参数统计——方法与应用》作为该课题的一个成果奉献给读者，以求这一统计方法能在我国市场研究、社会研究以及医学、企业管理等领域研究及应用方面更加普及。

由于时间仓促、水平有限，书中不免有许多缺憾，恳请读者加以指正。

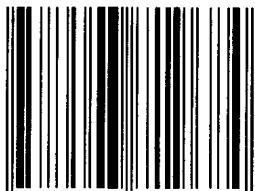
本书出版得到严建辉的支持，技术设计得到黄强、晏利东、范辉的大力帮助，作者在此一并表示感谢。

1994 年 11 月 30 日

责任编辑：吕 军

封面设计：崔葆进

ISBN 7-5037-1968-0



9 787503 719684 >

定价：14.80 元

试读结束：需要全本请在线购买：www.ertongbook.com

目 录

第一章 导 言	(1)
一、测量的层次	(1)
二、统计检验	(4)
三、非参数统计方法	(6)
第二章 单样本非参数检验	(10)
§ 2.1 χ^2 检验	(10)
§ 2.2 Kolmogorov-Smirnov 检验	(17)
§ 2.3 符号检验	(23)
§ 2.4 Wilcoxon 符号秩检验	(40)
§ 2.5 游程检验	(47)
第三章 两个相关样本的非参数检验	(56)
§ 3.1 符号检验	(57)
§ 3.2 Wilcoxon 符号秩检验	(60)
第四章 两个独立样本的非参数检验	(63)
§ 4.1 Mann-Whitney-Wilcoxon 检验	(63)
§ 4.2 Wald-Wolfowitz 游程检验	(70)
§ 4.3 两样本的 χ^2 检验	(73)
§ 4.4 两样本的 Kolmogorov-Smirnov 检验	(77)
第五章 k 个相关样本的非参数检验	(84)
§ 5.1 Cochran Q 检验	(84)
§ 5.2 Friedman 检验	(89)

第六章	k 个独立样本的非参数检验	(95)
§ 6.1	Kruskal-Wallis 检验	(95)
§ 6.2	k 个样本的 χ^2 检验	(103)
第七章	两个样本的相关分析	(109)
§ 7.1	等级相关	(109)
§ 7.2	Kendall 秩相关	(118)
§ 7.3	偏秩相关	(127)
第八章	k 个样本的相关分析	(131)
§ 8.1	Kendall 完全秩评定协和系数	(131)
§ 8.2	Kendall 不完全秩评定协和系数	(138)
§ 8.3	Friedman 检验和多重比较	(144)
第九章	列联表中的相关测量	(150)
§ 9.1	列联表相关测量的有关问题	(150)
§ 9.2	列联表的 χ^2 检验及相关测量	(156)
§ 9.3	列联表的 PRE 测量法	(163)
第十章	对数线性模型	(196)
§ 10.1	高维列联表	(196)
§ 10.2	对数线性模型的类型和参数估计	(206)
§ 10.3	模型的检验和选择	(221)
	参考文献	(233)
附 表		(234)
附表 I:	χ^2 分布表	(234)
附表 II:	二项分布表	(236)
附表 III:	单样本 $K-S$ 检验统计量	(248)
附表 IV:	正态分布表	(249)
附表 V:	标准正态分布表	(250)

附表 VI:	带有 $Q = 0.05$ 的累积二项分布表 (用于符号检验)	(252)
附表 VII:	Wilcoxon 符号秩检验统计量	(254)
附表 VIII:	游程分布的数目	(258)
附表 IX:	上下游程分布的数目	(263)
附表 X:	Mann-Whitney-Wilcoxon 分布表	(266)
附表 XI:	两样本 $K-S$ 检验统计量	(277)
附表 XII:	$K-W$ 检验统计量	(280)
附表 XIII:	Spearman 等级相关统计量	(282)
附表 XIV:	Kendall τ 统计量	(285)
附表 XV:	Kendall 协和系数检验的统计量	(287)
附表 XVI:	多重比较的临界值 Z	(289)
附表 XVII:	F 分布表	(290)

第一章 导 言

一、测量的层次

所谓测量,就是根据一定的法则,给事物或事件分配一定的数字或符号。如测量职工对所从事职业的满意程度,可以分配0至10的数字,不满意给0分,很满意给10分,介于二者之间的分配中间的数字。于是,0至10这些抽象的数字符号就依据上述规定的法则,表示职工对所从事职业的满意程度。测量的作用在于准确地描述事物的类型、性质、状态,同时对事物之间的差异进行准确测量和比较。事物只有通过测量,才有可能选择适当的统计模型或公式,进行一系列的统计分析。

目前,广泛采用的测量层次为四种类型的尺度:定类尺度、定序尺度、定距尺度和定比尺度。

(一) 定类尺度

定类尺度是按照事物的某些特征辨别和划分它们异同的一种测量层次,也被称作类别尺度、名义尺度。如性别、职业、民族等,都是按照事物的性质、类别区分的,均属定类尺度。

测量时,无论哪一种测量层次,都必须具有完备性和互斥性。完备性,是指用这种尺度测量某事物时,必须对这一事物所包括的各种情况都能进行测量。如测量性别时,应包括性别的各种情况:男、女,不能有遗漏。互斥性,是指用这种尺度测量时,不能有任一被测量对象跨越类别,即事物的各种情况具有互相排斥的不同值。如测量性别时,分为互斥的两类:男、女,某人或属于男,或属于女,不能既属于男又属于女。测量层次的完备性和互斥性,可以保证测量的准确无误。

定类尺度只能将事物分类,不能用以反映事物的数量状况,有

时,为了识别不同的类别,也用一定的数字和符号表示某类事物。如职工对所从事的职业是否满意,可以用“0”表示不满意。用“1”表示满意,这仅仅是人们赋予的识别标志,并不说明事物的数量。定类尺度是最低一个层次的测量尺度,它不能进行算术运算,而只能进行“=”或“≠”的逻辑运算。定类测量数据的描述性统计量有:众数、频数等。

(二) 定序尺度

定序尺度是按照事物的某种特征依顺序和级别进行排列的一种测量层次,也称作顺序尺度、等级尺度。例如,测量职工的文化程度可以采用定序尺度,分为:大专以上、中专或高中、初中、小学,显然前面的类别要比后面的类别高,即前面的文化程度最高,依次降低。定序尺度不仅能够区分事物,即对事物进行分类,而且可以反映事物在高低、大小、强弱上的差异,也就是使类别之间具有次序比较关系。定序尺度是比定类尺度高一层次的测量,它不仅能进行“=”或“≠”的运算,还能进行“>”、“<”的运算。最适合描述定序尺度中数据集中趋势的统计量是中位数,反映离散程度的是分位数。

(三) 定距尺度

定距尺度是不仅能将事物区分类别和等级,而且可以确定其之间的数量差别、间隔距离的一种测量层次,也称作间隔尺度、区间尺度。例如,对学生学习成绩的测量,甲为90分,乙为85分,甲乙学生成绩间距为 $90 - 85 = 5$ 分,这一测量就是定距尺度。定距尺度没有绝对的零点,也就是说,在这种测量中,任何两个间隔的差异与零点无关。例如,某门课程成绩的百分制测量,0分不表示某考生没有这门课方面的知识, $90 - 85 = 5$ 分,只表明甲生比乙生在这门课考试成绩中多5分,并不能说明甲生掌握的这方面知识是乙生的大约1.06倍($90/85$)。定距尺度在实际应用中较为普遍,象温度、智商等都是定距测量。定距尺度是一种定量的测量层次,它不仅能反映事物的类别和顺序,而且能反映事物的具体数量

和数量之间的距离。它是比定序尺度又高一层次的测量,不仅能进行“=”、“ \neq ”,“ $>$ ”、“ $<$ ”的运算,还能进行“+”、“-”的运算。定距尺度中描述性统计量,除了反映集中趋势的众数、中位数、均值外,还有反映离散程度的方差、标准差等,一般的定量统计方法都可以在这一测量层次应用。

(四) 定比尺度

定比尺度是在定距尺度上增加绝对零点的一种测量层次,也称作等比尺度、比率尺度。例如对职工年龄的测量,这里0岁是非任意的,一个人年龄不可能比0岁更小,这一测量尺度对所有人都一样。若甲为40岁,乙为20岁,则甲的年龄是乙的2倍,这就是定比尺度的测量。是否具有实际意义的零点存在,是定比尺度与定距尺度的唯一区别。定比尺度由于有一绝对零点存在,因而比定距尺度更利于反映事物之间的比例或比率关系,它是所有测量层次中最高一层的测量,不仅能进行“=”、“ \neq ”,“ $>$ ”、“ $<$ ”,“+”、“-”的运算,而且能进行“ \times ”、“ \div ”的运算。在定比测量中,描述性统计量不仅有算术平均的均值,还有几何平均的均值,不仅有方差、均方差,还有变异系数等。

(五) 四种测量尺度的关系

四种测量尺度有着不同的特点,其主要表现在作用和运算性质上,但它们之间又有较为密切的关系。首先,这些测量尺度之间有着包含关系,即高一层次的测量尺度总是包含低层次的测量尺度。定序尺度包含了定类尺度所有运算性质,定距尺度包含了定序、定类尺度所有运算性质,而定比尺度则包含了所有测量层次的运算性质。其次,四种测量尺度之间,低级的测量尺度往往能用较高级的测量尺度形式表示。例如,对学生考试成绩的测量,进行定类测量可分为及格、不及格;若将及格的成绩高低排序,可分为优、良、中、及格,这是定序尺度;若再将各顺序级给出等级分,则按百分制测量,优:90分以上,良:80—90分,中:70—80分,及格:60—70分。对同一事物的测量可以用多种尺度时,为避免信息的

丢失,应尽量将低层次测量尺度变成较高层次的尺度来测量。由于不同的测量层次具有不同的数学性质,因而在统计资料的收集、整理、分析过程中,往往需要采用不同的统计方法,也就是说,统计方法的运用总是与所选择的测量尺度相联系。在实际应用时,要首先弄清统计方法适用的测量尺度与所获得的资料采用的测量尺度是否一致。

二、统计检验

近代统计学的中心课题是统计推断。在统计推断中涉及这样的问题:如何利用部分事件的观察作出大量事件的结论。例如,要确定几种牌号的彩色电视机在我国居民中哪种最受欢迎,可以这样去搜集资料:到一家最大的商场站在柜台边,计数一天中每种牌号彩电的销售数量,几乎可以肯定那几种牌号彩电销量不同。但能否推断:那一天在这家商场销量最多的彩电是最受我国居民欢迎的呢?这取决于那种彩电的销售地域,也取决于那家商场的代表性,还取决于所观察的那些买主的代表性。统计检验正是要解决这个问题:如何根据样本值判断所得出的结论是否正确。

统计检验的一般步骤为:

陈述零假设(H_0);

选择一种统计检验来检验 H_0 ;

给定显著性水平 α 和样本容量 n ;

求出 H_0 成立时统计量的抽样分布;

确定否定域;

利用样本资料计算统计检验值。若其值落在否定域,则拒绝 H_0 ;若落在否定域之外,则在所选择的显著性水平上,不能拒绝 H_0 。

(一) 零假设

零假设是一种无差别假设,表示要被拒绝的目的,也称原假设。备择假设(H_1)亦称对立假设,是与 H_0 相反的结论。若 H_0 被拒

绝, H_1 就可能被接受。例如, 研究两种药物对治疗同一种病的效果不同。这个结论是要研究的假设, 为了检验它, 一般把它陈述为备择假设, 可以用运算的形式写出为, $H_1: \mu_1 \neq \mu_2$, H_0 则是: $\mu_1 = \mu_2$, 表明两种药物对治疗同一种病的效果相同。如果搜集的资料允许拒绝 H_0 , 则 H_1 被接受, 这就支持了研究假设以及由它导出的理论。

H_1 的叙述是由研究假设的性质确定的。若研究假设只是考察两个事物有差异, 则备择假设 H_1 为 $\mu_1 \neq \mu_2$; 若考察其差值的方向, 则 H_1 或者为 $\mu_1 > \mu_2$, 或者为 $\mu_1 < \mu_2$ 。第一种检验为双尾检验(双侧检验), 后两种检验为单尾检验(单侧检验)。

(二) 显著性水平

假设检验是根据人们一条普遍的经验作为原则的, 即小概率事件在一次实验中很难发生。如果一旦发生, 就认为原来的假设不成立, 从而拒绝 H_0 。但是, 很难发生并不等于决不发生, 因此, 在得出对 H_0 的判定时, 可能会发生两类错误: 第一类错误是当 H_0 实际上为真时拒绝 H_0 ; 第二类错误是当 H_0 实际为假时接受 H_0 。第一类错误是“以真为假”的错误, 犯第一类错误的概率由 α 给出, α 越大, H_0 越容易错误地被拒绝; 第二类错误是“以假当真”的错误, 犯第二类错误的概率通常用 β 表示。由于犯这两类错误的概率之间存在着反比关系, 因而, 当样本数目 n 一定时, α 减小将使 β 增大。若希望同时减小犯两类错误的可能性, 必须增加样本数目 n 。

实际应用时, 人们通常只能控制犯第一类错误的概率, 也就是错误地拒绝 H_0 的概率, 这个概率就叫做显著性水平。它一般在进行统计检验时事先给定。在选定 α 的大小时, 应根据实际情况考虑。若宁可“以真为假”, 则应把 α 取得小些, 如 0.01; 否则, α 可取大些。一般检验时, 取 $\alpha = 0.05$, $\alpha = 0.01$ 较多。为了保证 β 不致太大, 样本数目不宜太少(如至少不小于 5)。

(三) 否定域

拒绝零假设 H_0 的区域称为否定域或拒绝域。否定域的大小

与显著性水平 α 的选取有关。对于同一组样本数据, 选取不同的 α 值, 可能得到截然相反的结论, 如取 $\alpha = 0.05$, 可能拒绝 H_0 , 而取 $\alpha = 0.01$, 则不能拒绝 H_0 。选择 α 的原则, 不愿“以假当真”, 而宁可“以真为假”时, 应取很小的 α 值, 正是基于否定域与 α 的关系建立的。

否定域的位置(不是大小)与备择假设 H_1 的性质有关。若 H_1 是指出预定方向的, 如 $H_1: \mu > \mu_0$, 则假设检验为单侧检验(单尾检验); 若 H_1 未指出预定方向, 如 $H_1: \mu \neq \mu_0$, 则为双侧检验(双尾检验)。图 1.1 是 $\alpha = 0.05$ 的单侧检验否定域, 图 1.2 是 $\alpha = 0.05$ 的双侧检验否定域。可以看出, 对于同一显著性水平 α , 两种否定域的位置不同, 但总的大小并没有什么不同。

在进行统计检验时, 若根据样本数据计算的统计量数值落入否定域, 则认为零假设 H_0 不成立, 称作在显著性水平 α 下拒绝 H_0 ; 否则认为零假设 H_0 成立, 称作在显著性水平 α 下不能拒绝 H_0 。

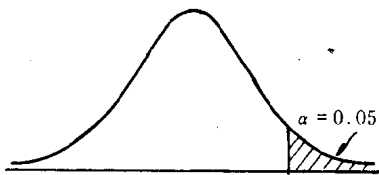


图 1.1 $\alpha = 0.05$ 的单侧检验
否定域

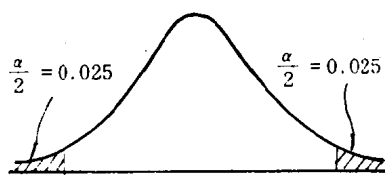


图 1.2 $\alpha = 0.05$ 的双侧检验
否定域

三、非参数统计方法

(一) 参数统计和非参数统计

在数理统计学中, 统计检验的种类很多, 而每一种统计检验都

与一种模型和一种测量要求相联系,只有在一定的条件下,某种统计检验才是有效的,而模型和测量要求则具体规定了那些条件。对那些其总体分布族或称统计模型只依赖于有限个实参数的问题,通称为“参数统计问题”,也就是说,总体分布服从正态分布或总体分布已知条件下的统计检验,称为参数检验,研究这一问题的统计分支称为参数统计。参数统计的大部分方法要求所分析的数据至少是定距尺度测量的结果。如统计学中的 t 检验、 F 检验等,都属于参数检验。

表 1—1 四种测量层次及适宜的统计量

测量层次	数学性质	描述统计量	适合的统计检验
定类尺度	$=$ 、 \neq	众数 频率 列联系数	非参数 检验
定序尺度	$=$ 、 \neq 、 $>$ 、 $<$	中位数 百分位数 Kendall τ Spearman r_s	
定距尺度	$=$ 、 \neq $>$ 、 $<$ $+$ 、 $-$	平均值 方差 Pearson r	参数检 验与非参 数检验
定比尺度	$=$ 、 \neq $>$ 、 $<$ $+$ 、 $-$ \times 、 \div	几何均值 标准差系数 多重相关系数	

当总体分布不能由有限个实参数所刻划时的统计检验,称为非参数检验,也就是说,统计检验的正确、有效并不依赖于总体的一个特定的统计模型即并不取决于总体分布时,称为非参数检验。

非参数检验通常认为是总体分布不要求遵从正态分布或总体分布未知条件下的统计检验,研究这一问题的分支称为非参数统计。非参数统计方法可以适用低于定距尺度测量的数据。

目前,关于什么样的统计技术准确地说是非参数统计技术并没有一个普遍性的一致认识^①,上面的论述是本书所用的定义范围。

从上面的论述可以看到,非参数检验的假定条件要比参数检验宽松得多,不仅对总体分布,而且对数据的测量层次。因而,其应用的范围要比参数检验广泛。表 1—1 概括了各种测量层次的数学性质及常用的描述统计量和统计检验种类。

(二) 非参数统计的优点

非参数统计是相对于参数统计而出现的,其优点也应在与参数统计的对比中考察。

1. 适用面广。从表 1—1 可以看出,非参数统计方法的适用面较参数统计广。它不仅可以用于定距、定比尺度的数据,进行定量资料的分析研究,还可以用于定类、定序尺度的数据,对定性资料进行统计分析研究。如利用问卷调查资料,进行居民对某几种商品质量满意程度是否相等的分析研究;利用民意测验,分析研究居民对几种房改方案的支持率是否有差异等等。而这些方面的研究是参数统计方法所不能及的,只有应用非参数统计方法。

2. 假定条件较少。经典的参数统计要求被分析的数据的总体遵从正态分布,或至少要遵从某一特定分布且为已知。而非参数统计假设条件比较少,并不要求总体分布遵从什么具体形式,有时甚至不需要什么假定,因此更适合一般的情况。

3. 具有稳健性。稳健性(Robustness)反映这样一种性质:当真实模型与假定的理论模型有不大的偏离时,统计方法仍能维持较为良好的性质,至少不会变得很坏。参数统计方法是建立在严格

^①参见参考文献 [5]。

假设条件基础上,一旦假定条件不符合,其推断的正确性就会不存在。非参数统计方法由于都是带有最弱的假设,对模型的限制很少,因而天然地具有稳健性。这是非参数统计方法常被使用的一个理由。

(三) 非参数统计的缺点

当定距或定比尺度测量的数据能够满足参数统计的所有假设时,非参数统计方法虽然也可以使用,但效果远不如参数统计方法。这时,如果要采用非参数统计方法,唯一可以补救的办法就是增大样本容量,用大样本,弥补由于采用非参数统计方法而带来的损失。譬如说,通过 90 次独立观察获取的数据足以保证参数统计所要达到的精度,而若用非参数统计方法,可能至少需要 100 次独立地观察以获取数据。

在一些定距或定比尺度测量的数据中可以广泛地应用的参数统计方法并没有与其相对应的非参数统计方法(从目前研究的成果还没有发现)。由于参数统计方法对数据有较强的假定条件,因而当数据满足这些条件时,参数统计方法能够从其中广泛地充分地提取有关信息。非参数统计方法对数据的限制较为宽松,因而只能从其中提取一般的信息。当数据资料允许使用参数统计方法时,采用非参数统计方法会浪费信息。