颜金锐 著

科研中常用的统计方法

自由分布统计检验

中国统计出版社
China Statistics Press

科研中常用的统计方法

——自由分布统计检验

颜金锐 著

中国统计出版社
China Statistics Press

(京)新登字041号

图书在版编目(CIP)数据

科研中常用的统计方法：自由分布统计检验/颜金锐著.
-北京：中国统计出版社，2002.2
ISBN 7-5037-3707-7

Ⅰ.科…
Ⅱ.颜…
Ⅲ.统计分布
Ⅳ.0212.2

中国版本图书馆CIP数据核字(2002)第001465号

作　　者/颜金锐
责任编辑/蔡启新
封面设计/张　冰
出版发行/中国统计出版社
通信地址/北京市西城区月坛南街75号　邮政编码/100826
办公地址/北京市丰台区西三环南路甲6号
电　　话/(010)63459084、63266600-22500(发行部)
印　　刷/科伦克三莱印务(北京)有限公司
经　　销/新华书店
开　　本/850×1168mm　1/32
字　　数/152千字
印　　张/6.25
印　　数/1-1500册
版　　别/2002年7月第1版
版　　次/2002年7月第1次印刷
书　　号/ISBN 7-5037-3707-7/O·42
定　　价/18.00元

序　　言

自由分布统计检验是推断统计中假设检验的一个分支。由于它可以适用于各种类型的统计数据分析，所以，尽管它的起步较晚，是上个世纪 50 年代以后才发展起来的，但现已广泛应用于医药疗效分析、农业田间试验分析、民意调查分析、市场调查分析、股票市场股价分析以及体育运动与文艺活动各种成绩分析等诸多领域，成为常用的科学研究统计分析方法。它是统计科学一个非常值得深入研究和推广的学科。

本书是通过多年对大学本科和研究生教学实践后整理形成的。现交付出版，旨在为丰富和发展我国的统计科学添砖加瓦。无论从事何种行业的理论工作者和实际工作者，凡是需要利用统计资料进行分析、比较、评价的，本书可以作为一个有力的助手。因此也可以说，将本书定义为工具书。本书可作教材使用，也可作自学读物。

为了读者应用的方便，本书以简炼通俗的语言表达各种检验的原理和方法，选择典型的实例，详细列出分析的步骤，每章末了还设计若干练习题，并在书后附上参考答案。

感谢已毕业的研究生肖玉梅同志将本书草稿打印成

册，也感谢我的妻子高级统计师郑清莹同志协助我整理陈年书稿及进行校对所付出的辛勤劳动。

本书写作过程中，参阅了许多版本的同类型书籍，并采用了其中一些材料及附表，谨在此向有关作者致以诚挚的谢意。

限于本人水平，书中恐有不少缺点或错误，欢迎读者批评指正。

颜金锐

2002年1月29日

于厦门大学

目　　录

第一章　导论

第二章　检验功效和功效效率

第三章　列名尺度数据:单样本场合

第四章　列名尺度数据:独立样本

第五章　列名尺度数据:相关样本

第六章　顺序尺度数据:单样本场合

第七章　顺序尺度数据:多个独立样本

第八章　顺序尺度数据:相关样本

第九章 间隔尺度数据和比率尺度数据
——数据的随机性检验

练习题参考答案 …………………………… (120)

附表

第一章 导 论

本章要点：

1. 了解自由分布统计检验的由来；
2. 了解概率的表达方法；
3. 了解如何利用二项分布计算各种事件的概率；
4. 了解单侧概率值和双侧概率值的区别；
5. 熟悉抽样分布的概念；
6. 了解如何确立二分类总体的抽样分布；
7. 了解如何利用抽样分布计算各种事件的概率。

自由分布统计检验是统计学的一个分科

统计学是一门关于数据资料的收集、整理、表述和分析的科学。统计方法是自然科学、社会经济、工程技术等各研究领域与工作部门必要的基本的数量分析手段，是从大量数据资料中提取主要的有用信息的有力工具。统计学可以区分为两大部分，其一为描述统计，它主要是探讨获得整个总体的全部数据之后，如何通过有关的“量”来有效地概述、刻划或表示，使整个总体的特征明显地表现出来，而这些“量”通常正是用来测定和描述所收集的特定数据资料的；另一部分则是推断统计，它主要是探讨如何在对样本进行观测而获得数据资料的基础上，用样本已知的有关的“量”来对整个总体相应的“量”加以推断估计而作出结论。随着社会的发展和科学技术的进步，推断统计日益显示出它的重要性，在任何科学的研究中，要寻求一般的普遍规律都离不开它。描述统计与推断统计的区别并不在于所收集

的数据资料，也不在于所作的计算，而是在于数据资料的用途，及其使用的目的上。二者常可结合使用，而且后者往往须以前者所得结果为依托。

上个世纪五十年代以前，在统计学的发展中，最先出现的推断统计方法都对样本所属总体的性质作出若干假设，即对总体的分布形状加以某些限定。例如，假设样本所由出的总体必须是正态分布，或者假设两个样本取自具有相同方差的总体，等等。这种对总体分布形状等加以某些限定的推断统计方法，就称为限定分布统计方法(Distribution - Specified Statistical Methods)。由于这种方法所要推断的总体特征值(如平均值、方差等)是一些未知的“参数”，故在一些书本上其称为参数统计，其所做的假设检验则称为参数检验。这种对所要推断的总体分布形状等事先作出某些限定或假设，在应用上存在一定的局限性，它只有在关于总体分布的假设成立时，所得出的结论才被认为是正确的。因此，它在很多场合不便使用。

基于对上述存在问题的考虑，自上个世纪五十年代以来，一些杰出的统计学家相继发展了许多对总体不作太多的或严格限定的推断统计方法。由这些方法引出的结论所需的限定较少，因此，为了同上述的参数检验区别起见，有些书本上就将其称为非参数统计(Nonparametric Statistics)，其所做的假设检验则称为非参数检验(Nonparametric test)。事实上，“非参数”这一名称，同方法本身不相称，容易使人们误解为没有参数或不含参数。而如果不含参数或者没有参数，也就没有推断可言。因此，由于这种统计推断方法的特点是不依赖于总体分布的具体形状，仍能对总体作出推断，所以，我们称这种统计检验为自由分布统计检验(Distribution - Free Statistical test)，或自由分布统计方法(Distribution - Free Statistical Methods)。

严格说来，非参数检验同自由分布检验是有区别的。前者不陈述关于参数特定值的假设，而后者对样本所由出的总体分布具体形状不作假设。然而，自由分布统计比之非参数统计更能表达所做检验的性质。这里所谓的“自由”，是相对的。例如，在一个自由分布统

计中，有时假定总体具有连续性甚至是对称分布，但对分布的具体形状是不限定的。在一切场合，所检验的统计量分布必须是已知的，但是对于自由分布统计来说，这种分布是基于样本的特性，而不是基于总体的特性。

概率与概率分布

一、概率的表达方法

概率是用频率来表达的。频率变动范围在 0.00 至 1.00 之间。当 $p=0.00$ 时，表明事件出现的概率为 0；反之，当 $p=1.00$ 时，表明事件必然出现。

例：$p=0.50$，表明在 100 次机会中，事件将出现 50 次

$p=0.95$，表明在 100 次机会中，事件将出现 95 次

$p=0.25$，表明在 100 次机会中，事件将出现 25 次（或 4 次机会中出现 1 次）

概率可以用有利于给定结果的事件数同所有可能事件总数对比而得的频率来表示，即：

$$P_A=\frac{\text{有利于 }A\text{ 的事件数}}{\text{有利于 }A\text{ 的事件数}+\text{不利于 }A\text{ 的事件数}}$$

例：若定义一枚硬币的“面”为事件 A，硬币的“背面”为事件 $\bar{A}$，则投掷硬币一次而出现“面”的概率是：

$$P_A=1/(1+1)=1/2$$

若定义一颗骰子的“2”为事件 A，骰子的“1，3，4，5，6”为事件 $\bar{A}$，则投掷骰子一次而出现“2”的概率是：

$$P_A=1/(5+1)=1/6$$

二、二分类事件例解

研究者探讨的许多统计总体是由二项式组成的，例如，是与非，

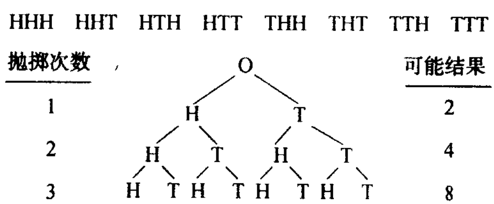
男性与女性,面与背,正确与错误,赞成与反对,等等。这种统计总体通常就称为二分类母体。二分类母体进行试验所获得的可能结果通常可以一一计数,并可据以计算各种事件的概率。

例:假定你在着手一项掷硬币的试验。倘若一枚硬币抛掷一次,有两种可能结果:面朝上或背朝上。每种结果多半是相等的。于是,

$$P_H(\text{面朝上的概率}) = P_T(\text{背朝上的概率}) = 1/2$$

倘若一枚硬币抛掷二次,就有四种可能结果:两个面朝上,一个面和一个背朝上,一个背和一个面朝上,两个背朝上。如果我们不计较获得面朝上和背朝上的先后顺序的话,各种可能结果就可以归纳为:一条途径获得两个面朝上,两条途径获得一个面朝上和一个背朝上,一条途径获得两个背朝上。

倘若一枚硬币在第一次抛掷时出现面朝上,它在第二次抛掷时可能出现面朝上或者出现背朝上;倘若第二次抛掷再出现面朝上,它在第三次抛掷时仍可能出现面朝上或者出现背朝上。从下列分枝图中,可以枚举出八种可能结果。



现在,可以对这些可能结果编制一张频数分布表如下:

一枚硬币抛掷三次,$n=3$

可能结果 x	出现次数 f
三个面 3	1
二面一背 2	3
一背二面 1	3
三个背 0	1
	8

每一种可能结果的频数除以可能结果总数获得概率分布如下：

x	f	P_x
3	1	0.125
2	3	0.375
1	3	0.375
0	1	0.125
	8	$\sum Px = 1.000$

有了概率分布，就可以提出关于各种不同结果的概率问题。

例 1：获得两个及以上面朝上的概率是多少？

$$P_{x\geqslant 2} = P_{x=2} + P_{x=3} = 0.375 + 0.125 = 0.50$$

例 2：获得三个及以上面朝上的概率是多少？

$$P_{x\geqslant 3} = P_{x=3} = 0.125$$

三、单侧和双侧的概率值

上述例举中，只着眼于从分布的一个终端获得概率值。诸如提出这样的问题：一枚硬币抛掷三次而获得两个及以上面朝上的概率是多少。而在推断统计学中，我们时常需要用一个分布的两个终端的形式来表示概率，诸如获得结果为如同某个特定值那样少有事件的概率是多少？

例 1：一枚硬币抛掷三次，获得结果为：三个面朝上或者同样少有事件的概率是多少？由于三个背朝上和三个面朝上同样是不常出现的事件，所以，将分布的两端的概率值加总就得出问题的答案。

$$P_{x=3\text{或}x=0} = P_{x=3} + P_{x=0} = 0.125 + 0.125 = 0.250$$

例 2：在一对骰子抛掷一次中，获得其点数之和为 3 及以下或同样不常出现事件的概率是多少？由于点数之和等于 11 及 12 与点数之和等于 3 及 2 同样是不常出现的，这两个极端的概率值为：

$$\begin{aligned} P_{x\leqslant 3\text{或}x\geqslant 11} &= P_{x=3} + P_{x=2} + P_{x=11} + P_{x=12} \\ &= 0.0556 + 0.0278 + 0.0556 + 0.0278 = 0.1668 \end{aligned}$$

抽样分布

抽样分布是推断统计学的关键概念之一。倘若从一个既定的母体中按照固定的样本容量抽取全部可能的样本，抽样分布就是某些样本统计量可能值在理论上的概率分布。

抽样分布提供理论上的概率值。以这个理论上的概率值为背景，我们就可以评价关于各种样本统计量的陈述。当我们基于一枚硬币抛掷三次获得“面朝上”的概率分布时，实际上就是编制了一个二项变量在 $n=3$ 时的抽样分布。因而，我们就可以描述这个分布的各种结果（诸如一个面朝上，二个面朝上，等等）的概率。

下表列示的是在 $n=12, P=Q=1/2$ 时，二项变量的抽样分布：

x	f	Px
12	1	0.0002
11	12	0.0029
10	66	0.0161
9	220	0.0537
8	495	0.1209
7	792	0.1934
6	924	0.2256
5	792	0.1934
4	495	0.1209
3	220	0.0537
2	66	0.0161
1	12	0.0029
0	1	0.0002
	4096	1.000

现在，让我们将问题稍许改变一下。假定你想提出一项关于二项变量的试验结果，那么，你的任务就是判别是否可能来自上述所确立的抽样分布（即其中 $P=Q=1/2$）的母体。

为有助于作出决策，特规定如下规则：

（1）倘若问题中的事件（或更少有的事件）出现的次数等于或少

于 5%，我们将对这种结果判定为非来自一枚硬币抛掷 12 次的母体。这称为否定的决策。

(2)倘若问题中的事件(或更少有的事件)出现的次数大于 5%，我们将可以认为试验的结果是来自既定的母体。这称为未能否定的决策。(请注意，未能否定的决策并不意味着肯定断言事件或结果是来自既定的母体，这只不过是承认它的可能性相当大而已)。

例 1：从一枚硬币抛掷 12 次的母体中，是否可能出现等于或少于 3 个正面的结果？

解：$P_{x\leqslant 3} = P_{x=3} + P_{x=2} + P_{x=1} + P_{x=0}$

$= 0.0537 + 0.0161 + 0.0029 + 0.0002$

$= 0.0729$

由于 $P_{x\leqslant 3} = 0.0729 > 0.05$，所以，根据规则(2)，我们作出未能否定(即予以肯定)的决策。

例 2：从一枚硬币抛掷 12 次的总体中，是否可能出现等于或大于 11 个正面的结果？

解：　$P_{x\geqslant 11} = P_{x=11} + P_{x=12} = 0.0029 + 0.0002 = 0.0031$

由于 $P_{x\geqslant 11} = 0.0031 < 0.05$，所以根据规则(1)，我们作出否定的决策，即结论是从一枚硬币抛掷 12 次的母体($P = Q = 1/2$)而出现等于或大于 11 个正面的结果是不大可能的。

检验统计假设

一、统计检验

运用统计方法检验一个事先作出的假设，这一假设叫做统计假设。对这一假设进行的检验，就称为(统计)假设检验，或称为统计(假设)检验。

统计检验的基本思路是这样的：

(1)为了检验一个“假设”是否成立,先假设它是成立的,然后看看接受这个假设之后是否会导致不合理的结果。倘若结果是合理的,就接受它;倘若不合理,则否定原假设。

(2)所谓导致不合理结果,就是看是否在一次观察中出现了小概率事件。根据实际抽样推断原理,小概率事件在一次抽选中是不大可能出现的。如果一旦出现,就很不合常规。因此,当然要怀疑原假设的正确性,从而否定原假设。上述两个例子的说明,实际上已经体现了统计检验的这种思想。

二、显著性水准

在统计检验中,判别假设是否合理,是根据一定标准来确定的。这个标准是人们事先根据主观选定的概率值,用符号 α 表示。这个 α 值,通常称为显著性水准。

上述样本统计量的抽样分布使得人们有可能区分哪些事件是经常出现的,哪些事件是不常出现的。概率理论提供了作出这种评价的基础。因而,倘若我们对一个极少出现的事件作出结论(例如一次试验的结果),断言非随机因素(例如实验变量的影响)是这个不常出现事件的起因将是可靠的。

在上述两个实例中,我们实际上使用了约定俗成的表示非随机因素作用(即作出否定决策)的分界点。这个分界点就是 0.05 显著性水准。用 0.05 作出否定决策时,习惯上就把试验结果称为 0.05 水准的统计显著性。

当事件的出现等于或小于 5% 而实际上是随机波动造成的时候,我们也宁可用归纳的方法断言其结果归因于非随机因素的作用。

当事件由于随机因素影响其次数只出现 1% 及以下时,我们断言其结果归因于非随机因素,习惯上就称这一试验结果为 0.01 水准的统计显著性。

在统计检验中,显著性水准 α 究竟取多大为宜,并没有一个具体

的规定,通常是视研究对象的特点和要求的严格程度由决策者决定的,也就是说,α 一般是事先给定的。对于社会经济现象,α 取 0.05 一般就足够了,对于民意测验的检验有时取 $\alpha=0.10$,而对于质量要求严格的工程技术问题,则取小的 α 值,诸如 0.01 甚至 0.001 等等。

对于同一个问题,若 α 取值不同,可能会影响到检验的结果。因此,显著性水准一经确定,对统计检验就起着决定性的作用。

例 1:倘若研究者用 0.05 的显著性水准,而试验结果的概率是 0.03,那么,他就断言非随机因素在起作用。

例 2:对例 1 的试验结果,若研究者用 0.01 的显著性水准,他将不再断言非随机因素在起作用,而是作出未能否定的决策。

研究者作出否定决策所确定的显著性水准 α,由于它在统计检验中起着决策性的作用,故通常称它为检验水平。当 $\alpha=0.05$ 时,检验水平为 0.05;当 $\alpha=0.01$ 时,检验水平为 0.01。

三、虚无假设和对立假设

检验开始之前,研究者事先建立两种相互排斥的假设:一个是研究者通常希望它能被推翻的统计假设,但是没有较充分的依据就不轻易推翻,这是被检验的主体,称为虚无假设,用符号 H_0 表示;另一个是研究者希望它能够成立的统计假设,称为对立假设,用符号 H_1 表示,这是统计检验的根本目的所在。通常,如果研究的目的在于希望从样本观测值中对某一陈述取得强有力的支持,我们就将这一陈述的否定作为 H_0,而把陈述本身作为 H_1。这里,由于 H_0 是关于不存在差别的假设,所以,它一定含有一个等号在内。H_0 包含等号,能断言总体参数至少(或至多)等于某个特定值,这样,我们才有订立决策(接受或拒绝)的临界值的基础,才能决定当观察统计量是什么数值时,在规定的显著性水准下可以拒绝 H_0 或接受 H_0。

例 1:抛掷一枚硬币的试验。则

H_0:硬币是完善的,即 $P_H=P_T=1/2$