

工商业和经济学用 统计方法技术

第 三 册

校 内 用 书
注 意 保 存

中国人民大学统计研究室译

目 录

第十五章	非参数假设检验：两个以上样本	1
第十六章	统计质量控制	24
第十七章	简单回归和相关分析	57
第十八章	复回归和相关分析	90
第十九章	联系的非参数计量	130
第二十章	偶然不确定性下决策导论	152
第二十一章	贝叶斯定理和决策	175
附 录		193
附录 1	普通对数	
附录 2	对 数	
附录 3	平方和平方根	
附录 4	随机数表	
附录 5	t 的临界值	
附录 6	“威尔科克森”T 值	
附录 7	正态曲线的面积和纵坐标表	
附录 8	χ^2 的临界值	
附录 9	二项分布	
附录 10	控制图系数	
附录 11	“曼·惠特尼”检验：U 的临界值	
附录 12	“斯皮尔曼”等级相关系数：Rho 的临界值	
附录 13	“肯德尔”一致性系数，W；S 的临界值	
附录 14	F 分布；显著水平，5% 和 1% 的临界值	
附录 15	“柯尔莫哥洛夫、斯米尔诺夫两个样本的检验	
附录 16	（大样本，双尾检验）：D 的临界值	

第十五章 非参数假设检验： 两个以上样本

本章要介绍三种非常有用的非参数显著性检验，用以阐明研究两个以上样本的检验程序。这样多的样本通常称为 k 个样本。这些检验的基本目的是要判断样本结果相互之间是否有显著差别。换一个说法，设计这些检验为的是确定 k 个样本是否是从相同的全域（或同一的全域）中选出来的。

头两种检验——科其安 Q 检验和弗利德曼两向方差分析——研究的是：在某些方面有关系的多个随机样本。

科 其 安 Q 检 验

是由科其安在1950年设计的，他的 Q 检验只需要列名水平的资料 and 根据三个或三个以上的有关系的观察数列。为了说明这个程序，假设现有对管理人员在厂内专题训练的四种不同的方法，管理部门要求评价这四种方法的效果，这四种不同方法是电视，课程讲授，自学教材及综合方法。在训练规划开始之前，在公司范围内随机选择了20个负责管理的人员作试点调查，以便评价所提出的这四种方法的效果。

使用四种方法中某一种方法给试点调查小组讲一个专题，如对新工人的管理。在介绍结束时要求每一个管理人员对这种方法给以评价。每一个管理人员的回答被译成“1”（表示有效）或“0”（表示无效）。注意表15—1，只有少数几个管理人员给所有方法的评价都是有效的，而其他一些管理人员只认为某些方法是有效的，

或甚至认为没有一种方法是有效的。

表15—1 管理人员对介绍的四种方法的回答

管理人员号码	(1) 只用电视	(2) 只用讲授	(3) 自学教材	(4) 综合方法	L	L ²
1	0	0	1	1	2	4
2	1	0	0	1	2	4
3	0	1	0	0	1	1
4	0	0	0	0	0	0
5	0	0	0	1	1	1
6	0	0	1	1	2	4
7	0	1	0	0	1	1
8	0	0	0	0	0	0
9	1	0	0	1	2	4
10	1	1	1	1	4	16
11	0	0	1	1	2	4
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	1	1	0	1	3	9
15	0	0	0	1	1	1
16	0	0	1	1	2	4
17	0	1	0	1	2	4
18	0	0	0	1	1	1
19	0	0	1	0	1	1
20	0	0	1	1	2	4
总计	4	5	7	13	29	63
	ΣS_1	ΣS_2	ΣS_3	ΣS_4	ΣL	ΣL^2

这里，应注意20个管理人员中有4个认为电视是一种有效的工具，5个人认为讲授方法是有效的，如此等等。

科其安 Q 检验要解决的问题是：关于提出的这四种方法其效果有没有显著差别。因此，虚无假设是：关于提供的方法在有没有有效的反应之间并没有区别。 H_1 ，替换假设为： k 个样本在有没有有效方面，有着显著区别。 H_1 的这种申述表明：要用双尾检验。

合适的统计检验为科其安的 Q 检验，因为资料是列名水平的并有着两个以上有关系的样本（本题中所谓有关系是指每一个管理人

员按照他个人意志行动)。而管理人员的回答则可以划分为“有效”和“无效”两种。

暂时选定0.05水准。这意味着可能犯第一类错误,也就是说,当应该接受虚无假设却拒绝了的机会是5%。

Q的一般公式是:

$$Q = \frac{(k-1) [k \sum S_j^2 - (\sum S_j)^2]}{k (\sum L) - \sum L^2}$$

应用到这个问题, Q将为:

$$Q = \frac{(k-1) [k (\sum S_1^2 + \sum S_2^2 + \sum S_3^2 + \sum S_4^2) - (\sum L)^2]}{k (\sum L) - \sum L^2}$$

这里:

k 是样本数。本题中,它为不同教育方法的数目,这里是4。

$\sum S_1$ 是赞成电视方法(认为它有效)的回答数的总和。 $\sum S_2$ 是赞成讲授方法回答数的总和,如此等等。

L 是每个管理人员回答赞成(认为有效)数目的总和。例如,样本中第一个管理人员认为两种教育方法是有效的。回答中认为有效的全部总和数是 $\sum L$,或29。

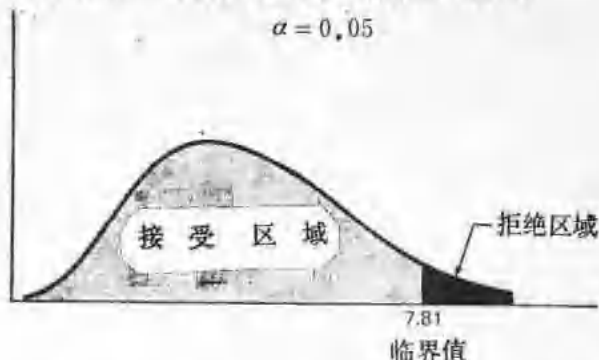
$\sum L^2$ 指的是:把每一个管理人员赞成某一方法的回答数(认为有效数)平方起来,然后求其总和, L^2 的总和是63。

Q的抽样分布接近于自由度 $k-1$ 情况下的 χ^2 。在这个题目中,有四个样本, $k-1=3$ 个自由度。图15—1用图解的方法来表示决策准则——如果计算出的Q小于临界值7.81,就接受虚无假设;但如果等于或大于7.81,则拒绝H。(临界值从附录8找出,当 $\alpha=0.05$,自由度为3。)

解Q,得到

$$\begin{aligned} Q &= \frac{(k-1) [k (\sum S_1^2 + \sum S_2^2 + \sum S_3^2 + \sum S_4^2) - (\sum L)^2]}{k (\sum L) - \sum L^2} \\ &= \frac{(4-1) [4 (4^2 + 5^2 + 7^2 + 13^2) - (29)^2]}{4 (29) - 63} \\ &= 11.04 \end{aligned}$$

图15—1 科其安Q的接受或拒绝的区域



计算出的Q值落在7.81外的区域，因而在0.05水准下拒绝了虚无假设，接受了替换假设 H_1 ，即对这四种教育方法之间的回答赞成的数目，有显著的差别。

必须注意到，科其安Q检验只是为了检验四种教育方法之间是否有总的差别。拒绝 H_0 而接受 H_1 ，并不表明，如综合方法最有效，它只是说明k个样本之间有着明显差别。如果有人要检验管理人员中赞成用综合方法的比例比如说要大于赞成用讲座的比例，则可以应用前章所讲到的t检验。

1. 全国知名的烤烘杯形糕点名为“要美”的烘焙厂又生产了三种新式糕点，它们看起来像“要美”，但味道与之不同。为了搜集一些初步资料，这个厂要求随机抽选出10个经常吃“要美”糕的人来尝试这三种糕点。首先，每个人尝一下新式糕点中的一种，称之为第一号糕点，然后说明他是否喜欢这种糕点超过“要美”糕。并用同样步骤尝试第二号，第三号糕点。回答为“是”或“不”（“是”表示这个人喜欢这种新糕点超过“要美”糕，“不”则表明他认为不比“要美”糕好）。

尝 试 者	糕 点		
	第一号	第二号	第三号
1	不	是	不
2	不	是	是
3	不	是	不
4	不	是	不
5	是	是	是
6	不	是	不
7	不	是	不
8	不	是	不
9	不	是	不
10	不	是	不

这个制造厂要求确定这个测试小组对这三种新产品的反应是否有总的差别。

- 规定虚无假设。
- 本题所提供的是何种计量水平的资料?
- 用0.10显著性水准和双尾检验, 阐述决策准则。
- 作出关于这三种糕点的决策。
- 这是否说明第二号糕点比“要美”糕好? 解释你的结论。

用本章末的答案检查你的答案

弗利德曼两向方差分析

回顾一下前一部分讲到科其安Q检验, 它是用来决定三个或更多的样本在某些特点方面有没有明显区别的。应用这种检验要假定观察材料为列名水平, 而且 k 个样本之间是有关系的 (通常指: 对每个问题的行动是由他自己的意志控制的)。

弗利德曼于二十世纪三十年代后期设计的显著性检验，则可应用于观察材料为顺序排列的，并有 k 个有关系的样本。举一个例子来说明：假设一个新的训练规划分成四个单元。每一个单元所列出的训练，采用一种不同的技术。参加这个项目训练的第一组包括有 14 个随机选出的雇员。在每一个单元结束时，对这 14 个雇员举行考试，并对成绩评给分数（见表 15—2）。

表 15—2 每种训练技术的考试成绩分数

工人姓名 (N=14)	技术 I	技术 II	技术 III	技术 IV
甲.....	20	6	9	15
乙.....	5	12	19	10
丙.....	11	21	8	16
丁.....	21	18	30	15
戊.....	8	12	20	16
己.....	9	7	10	12
庚.....	21	20	16	10
辛.....	18	27	9	12
壬.....	30	16	22	21
癸.....	22	27	19	18
子.....	10	8	4	12
丑.....	6	12	7	14
寅.....	10	12	21	20
卯.....	14	11	23	27

在继续往下说明这个问题之前，注意这个问题和前面应用科其安 Q 检验问题之间的相似之处。两个问题的目的都是一样的——确定四种技术或四种方法的效果，有无显著区别。然而，在前一问题中，关于管理人员的答案，只能分成“有效”或“无效”两种，也就是说资料为列名计量水平的。而这个问题之所以能应用弗利德曼检验，是因为成绩分数是按顺序排列的尺度，也就是说工人的成绩可以分组评级。

其虚无假设和替换假设是：

H_0 ：这四种技术总的效果之间并没有差别。

H_1 ：这四种技术总的效果之间有差别。

暂时选定0.05水准来检验虚无假设。

对于上述问题的统计检验应用弗利德曼双向方差分析，是合适的。因为（1）这里有 k 个样本；（2）样本是有关联的（每个雇员按他个人的意志，从事全部四个单元的训练）（3）成绩分数为按顺序的计量水平。

弗利德曼称之为 χ^2 的统计量接近于 $k-1$ 自由度的克方分布。因此，可以参看克方临界值表（附录 8）决定决策准则。克方的临界值是 7.81。像前面一样， k 代表样本数，这里有四个样本（四种技术）， $k-1=3$ 个自由度， $\alpha=0.05$ 。像通常一样，如果计算出来的 χ^2 小于临界值 7.81，就接受虚无假设，否则，就拒绝虚无假设。

为了应用弗利德曼检验，每个雇员的成绩分数要从低到高（或从高到低）顺序排列。用什么样的顺序无关紧要，但始终必须保持同样的程序。例如，表 15—2 用雇员甲的四个样本（四个训练单元）成绩分别是 20，6，9 及 15。最低分数 6 级别为 1，9 这个分数级别为 2，15 级别为 3，最高的分数 20 级别为 4。

分数从低到高顺序排列

雇员的名字 (N=14)	技术 I	技术 II	技术 III	技术 IV
甲.....	4	1	2	3
乙.....	1	3	4	2
丙.....	2	4	1	3
丁.....	3	2	4	1
戊.....	1	2	4	3
己.....	2	1	3	4
庚.....	4	3	2	1
辛.....	3	4	1	2

壬.....	4	1	8	2
癸.....	3	4	2	1
子.....	3	2	1	4
丑.....	1	3	2	4
寅.....	1	2	4	3
卯.....	2	1	3	4
总 计.....	34	33	36	37
	T_1	T_2	T_3	T_4

弗德利曼 χ^2 的一般公式是

$$\chi^2 = \frac{12}{Nk(k+1)} [T_1^2 + T_2^2 + T_3^2 + \dots + T_k^2] - [3N(k+1)]$$

应用到这个问题，它是：

$$\chi^2 = \frac{12}{Nk(k+1)} [T_1^2 + T_2^2 + T_3^2 + T_4^2] - [3N(k+1)]$$

这里：

N 是样本大小的总数，本题是14。

k 是有关系的样本数（栏），这里是4。

T_1 是一个样本（技术I）等级数的总和， T_2 是第二个样本（技术II）等级数的总和，如此等等。

将适当数值代入，我们得到：

$$\begin{aligned} \chi^2 &= \frac{12}{(14)(4)(4+1)} [(34)^2 + (33)^2 + (36)^2 + (37)^2] \\ &\quad - [3(14)(4+1)] = \frac{12}{280} [4,910] - 210 = 0.43 \end{aligned}$$

计算出来的 χ^2 值(0.43)小于克方临界值(7.81)，因此，在0.05水准下接受了虚无假设，即四种技术的全面效果没有显著性区别。

k 个样本的克方检验

由卡尔·皮尔逊在二十世纪初期创始的克方检验在前一章中是用于研究一个样本。如果有 2 个或 k 个独立样本，也可以用克方检验。由于计算的程序是一样的，只不过需要考虑到样本的数目是 k 个。

举一个例子来说明，假定我们要了解商业管理专业毕业生的学业成就（大学最后结业时的平均分数）和他们收入水平之间有没有关系。虚无假设为这两者没有关系，而 H_1 则申述为他们之间有关系。在这个检验中，暂时选定 0.05 显著性水准。

我们把工商业管理专业毕业生的学业成绩分成三类：平均水平以上，平均水平，及平均水平以下。这三类分别用最后结业时的分数 3.0—4.0，2.5—2.9999，和 2.0—2.4999 来表示（2 是能够毕业的必要的最低分数，而 4.0 表示一个毕业生在校学习成绩属于甲等）。这样，在理论上，就有了三个独立的样本。

把 751 个回答者的收入分成 4 种水平：低，中下，中上和高等。把这 751 个毕业生的回答编制成一个两向分类的表，叫作列联表（表 15—3）（这个表常常被看成是 3×4 列联表，因为它有 3 行 4 列）。

表 15—3 学业成就和收入水平

学 业 成 就	收 入 水 平				总 计
	低	中下	中上	高	
平均水平以上.....	22	31	31	8	92
平 均 水 平.....	67	80	73	17	237
平均水平以下.....	124	161	122	15	422

资料来源：R. D 梅森校友研究（俄亥俄州，托莱多，托莱多大学，工商业管理学院）

注意在表15—3之中，751个毕业生总数中有92个（12%）的分数在平均分数以上（3.0—4.0）。如果学业成就和收入之间毫无关系的话，就可以期望在收入水平低的213个毕业生中，有12%的人在校平均学业成绩的记录会在平均分数以上。这样，期望次数 f_e 这一格将是26（由 $213 \times 12\%$ 求出）。再进一步看，如果学业成绩与收入之间没有关系的话，中下收入水平的272个毕业生中也同样期望有12%的人有平均水平以上的分数。这一格的期望次数是33，由 $12\% \times 272$ 求出（见表15—4）。

用同样方法，再计算出另外一些格的期望次数。注意到在工商管理专业的751个毕业生总数中，有237个（占32%）在学校的分数属平均水平。如果学业成绩和收入之间没有关系的话，逻辑上在低收入这一级中也有32%的人分数属平均水平。这样这一格的期望次数是68（由 $213 \times 32\%$ 求出）。这样的程序继续下去，一直到算出所有各格的期望次数。表15—4列出了观察次数 f_o 和期望次数 f_e 。

表15—4 学业成绩与收入水平

学业成绩	收 入 水 平								总计	总计的
	低		中下		中上		高			
	f _o	f _e	f _o	f _e	f _o	f _e	f _o	f _e		
平均水平以上…	22	(26)	31	(33)	31	(27)	8	(5)	92	12
平 均 水 平…	67	(68)	80	(87)	73	(72)	17	(13)	237	32
平均水平以下…	124	(119)	161	(152)	122	(127)	15	(22)	422	56
总 计……	213	(213)	272	(272)	226	(226)	40	(40)	751	100

由于小数点进位的原故，每一列期望次数的总计可能不会刚好等于那一列观察次数的总计。如果发生了这种情况的话，就要通过调整1个或几个期望次数，使两个总计相等。

像前面一样，统计量 χ^2 通过下面的公式计算出来：

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

确定一个列联表自由度的一个比较方便的方法，是用（行数－1）×（列数－1）。在表15－4中，有3行4列的资料，因此自由度为（3－1）（4－1）＝6。参考克方表的临界值（附录8），提出所形成的决策准则。自由度为6， $\alpha=0.05$ 的临界值是12.59。如果算出来的 χ^2 值小于12.59，就接受 H_0 。（即虚无假设）；否则，就要拒绝 H_0 ，而接受替换假设 H_1 。

计算出来的克方是8.11（见下表的计算过程）。这个计算出来的值（8.11）小于临界值（12.59），因而落在接受域内。于是在5%水准下接受了虚无假设，即在校的学业成绩与收入水平并无关系。

平均等级分	收入	fo	fe	fo-fe	(fo-fe) ²	$\frac{(fo-fe)^2}{fe}$
平均水平以上……低		22	26	-4	16	$\frac{16}{26} = 0.62$
平均水平以上……中下		31	33	-2	4	$\frac{4}{33} = 0.12$
平均水平以上……中上		31	27	4	16	$\frac{16}{27} = 0.59$
平均水平以上……高		8	5	3	9	$\frac{9}{5} = 1.80$
平均水平……低		67	68	-1	1	$\frac{1}{68} = 0.01$
平均水平……中下		80	87	-7	49	$\frac{49}{87} = 0.56$
平均水平……中上		73	72	1	1	$\frac{1}{72} = 0.01$
平均水平……高		17	13	4	16	$\frac{16}{13} = 1.23$
平均水平以下……低		124	119	5	25	$\frac{25}{119} = 0.21$
平均水平以下……中下		161	152	9	81	$\frac{81}{152} = 0.53$
平均水平以下……中上		122	127	-5	25	$\frac{25}{127} = 0.20$

$$\begin{array}{rccccccc} \text{平均水平以下} \cdots \cdots & \text{高} & 15 & 22 & -7 & 49 & \frac{49}{22} = 2.23 \\ & & \hline & 751 & 751 & 0 & & & 8.11 \end{array}$$

这个问题的资料是按顺序排列的（因为收入和学业成绩均按从低到高依序排列）。不过，还必须注意到克方检验也可以应用于列名，间隔，及比率水平的资料。

前一章讨论过的关于应用克方检验的限制，这里要再重复一下。这个限制提到：如果期望次数小于5的格超过20%的话，就不能用这种检验。不过，可以像前面所建议的那样，把几类合并起来，以适合这种检验的要求。

2. 一个社会学家研究下面这样一个问题：“即一个人受的教育程度与其社会活动之间有没有关系？”她确定了三种教育程度：大学或完成大学学业；中学或完成中学学业；小学及以下。每个人有一份社会活动的纪录，例如：同一些人滚木球、跳舞、参加教堂集会等。社会学家按参加次数分类：超过平均水平，平均水平；低于平均水平。

教育程度	社 会 活 动		
	超 过 平均水平	平 均 水平	低 于 平均水平
大 学.....	8	2	5
中 学.....	8	5	7
小 学.....	4	3	13

- 上面的表叫什么表？
 - 规定虚无假设，在0.05水准下进行假设检验。
 - 注意某些格的次数小于5（为了计算方便）。不考虑限制性，计算克方。
 - 接受还是拒绝虚无假设，引用数字来证实你的决策。
 - 在这个问题中，特别说明了些什么？
- 用本章末的答案来检查你的答案。

简 要 概 述

非参数假设检验：两个以上样本

I 非参数假设检验：两个以上样本

A、目的。检查包括三个或三个以上样本的列名和顺序的计量水平某些假设申述的真实性。设计这些检验是用以确定 k 个样本是否选自相同全域（或同一的全域）。

II 科其安 Q 检验。

A、假定。需要列名的计量水平，样本是有关系的，这是指这个人按他自己的意志行动的；也就是说，他的回答包括在样本1，样本2，……之中。

B、用途。例如：要求20个人对推荐的具有中等价格水平汽车的五种试验模型，发表意见（喜欢或不喜欢）一般排列如下：

人	模 型				
	A	B	C	D	E
毕女士……………	喜欢	不喜欢	喜欢	喜欢	喜欢
西先生……………	不喜欢	不喜欢	不喜欢	不喜欢	喜欢

虚无假设是：对试验模型是否喜欢的反应，在五个样本间没有总的区别。

C、公式。把“喜欢”这个字换成 1，“不喜欢”换成 0，然后：

$$Q = \frac{(k-1) [k(\sum S_1^2 + \sum S_2^2 + \sum S_3^2 + \sum S_4^2 + \sum S_5^2) - (\sum L)^2]}{k(\sum L) - \sum L^2}$$

这里：

k 是样本数目

$\sum S_1$ 是对模型 A 回答喜欢的总数， $\sum S_2$ 是对模型 B 回答喜欢的总数，等等。

L 是每一个人回答喜欢的总数。

$\sum I_i^2$ 是每一个人回答喜欢数的平方，然后加总。

D、决策程序。Q的抽样分布接近于 $k-1$ 自由度的克方。参看附录 8 找出选定的显著性水准下的临界值。

III 弗利德曼双向的方差分析

A、假设。需要按顺序排列的资料和有联系的样本。

B、用途。上例中这 20 个人对五种模型的回答不是用“喜欢”或“不喜欢”表示，而是按等级排成 1, 2, 3 ……。

人	模 型 的 等 级				
	A	B	C	D	E
毕女士……………	2	3	1	4	5
西先生……………	5	4	1	3	2

虚无假设和替换假设是：

H_0 ：这 5 个试验模型的等级总的说来没有区别。

H_1 ：这 5 个试验模型的等级总的说来有区别。

C、公式：

$$\chi^2 = \frac{12}{Nk(k+1)} [T_1^2 + T_2^2 + \dots + T_k^2] - [3N(k+1)]$$

这里：

N 是样本中回答者的总数。

k 是有联系的样本数目（栏数）。

T_1 是其中一个样本（模型 A）等级数的总计。

T_2 是第二个样本等级数的总计等等。

D、决策程序。 χ^2 抽样分布接近于自由度为 $k-1$ 的克方。参看附录 8，在选定的显著性水准下找出临界值。

IV k 个样本的克方检验。

A、假定。至少是列名的计量水平，k 个独立样本。

B、用途。例：性别和收入之间没有关系。将资料分类整理编在一张列联表中。

性 别	收 入		
	低	平均	高
男			
女			

将各类观察次数和期望次数进行比较。

C、公式：

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

D、决策程序。利用通常的假设检验程序，计算 χ^2 。参看附录8，按选定的显著性水准找出临界值。

本章末的自我复习检查

两个问题都作，然后用章末答案来检查你所作的结果。
每题50分。

1. 一个十人专家小组对五种游览汽车的设计方案进行分等评定，给分从0（非常不满意）到100（杰出）。评定结果是：

专 家 号 码	设 计 方 案				
	A	B	C	D	E
1	40	90	10	46	18
2	60	8	98	12	9
3	2	80	90	98	99
4	82	60	5	50	55
5	42	99	3	2	60
6	88	3	99	40	45
7	8	50	9	10	82
8	14	3	29	19	10
9	9	10	5	7	1
10	12	8	7	6	50

用0.05水准来检验，说明总的等级没有显著的差别。