

秩和比法的应用

田凤调 著

$$RSR = \frac{\sum R}{m \cdot n}$$

$$RSR = \frac{\sum R}{m \cdot n}$$



人民卫生出版社

934

C81
T56

秩和比法

的

应用

田凤调 著

人民卫生出版社

图书在版编目 (CIP) 数据

秩和比法的应用/田凤调著. —北京:人民卫生出版社,2002

ISBN 7-117-05211-2

I. 秩... II. 田... III. 统计方法-应用
IV. C81

中国版本图书馆 CIP 数据核字 (2002) 第 092473 号

秩和比法的应用

著 者: 田 凤 调

出版发行: 人民卫生出版社 (中继线 67616688)

地 址: (100078) 北京市丰台区方庄芳群园 3 区 3 号楼

网 址: [http://www. pmph. com](http://www.pmph.com)

E-mail: [pmph @ pmph. com](mailto:pmph@pmph.com)

印 刷: 三河市富华印刷包装有限公司

经 销: 新华书店

开 本: 787 × 1092 1/16 印张: 24.5

字 数: 559 千字

版 次: 2002 年 12 月第 1 版 2002 年 12 月第 1 版第 1 次印刷

标准书号: ISBN 7-117-05211-2/R·5212

定 价: 47.50 元

著作权所有, 请勿擅自用本书制作各类出版物, 违者必究

(凡属质量问题请与本社发行部联系退换)

前 言

秩和比法是一种全新的广谱的实用数量方法，也称统计信息方法。此法集参数统计与非参数统计于一身，有描述有推断，能提高统计分析与再分析的水平，满足人们在统计研究与统计管理中的各种要求。

建立秩和比法的出发点：设计时既考虑到统计信息的“原材料”，又考虑到统计信息人员及使用者的现有水平。当今时代，我们拥有大量的统计资料，有些虽已经提取出统计信息，但在深度和广度方面还远远不能满足需要；而且有关资料信息也是层出不穷的，要求追补进来也是常事；加之，我们的原材料还含有“水分”，不那么经得起推敲，因而在应用精密统计分析方法处理时要谨慎行事。若是勉强套用，违背方法的立论假说，危险不言而喻。现有专业人员的水平有高有低，但大多数基层统计信息人员数理基础还较薄弱，先进的、高层次的数理统计与数学文献一时还难以掌握，因而不易正确使用。此外，数理统计方法使用的数学语言，立论严谨，一般有假设，有推导，限制多多。例如，如何营造一个真正的随机环境，什么是正态分布、方差齐性等等，实际取得的资料往往不是像书中所论述的那样。

秩和比法创建于1988年，此后立题研究，于1992年获中国预防医学科学院1992年科技三等奖（秩和比法为该课题的中心内容）。1993年由中国统计出版社出版专著《秩和比法及其应用》。用时三年（1993.5~1996.5）在卫生系统办班推广；用时七年（1996.5~2002.5）完成了《实用卫生统计学新编》与《秩和比法在缤纷世界中的应用》两个姊妹篇的撰写。本书为上述两稿的合订本，为秩和比法应用研究的全部。作为方法论的一种，作为一项有重要应用价值的复合信息载体，不但要应用在本专业的领域里，还要应用在各个专业中。

作者一贯主张，在方法论方面，除古典的概率论与数理统计外，社会经济统计、其他数量方法（如模糊数学、灰色理论、系统论、运筹学、软科学以及相关学科的某些数量方法等）均可引入。它们互相借鉴，互相渗透，互相补充，互相印证，从而可以增强统计信息生成的能力。我们要学会各种方法的沟通与链接，秩和比法更要强调这一点。

我们提倡统计信息生成方法的简易性和有效性。我们的目标是去粗取精，变繁杂为精纯，辩偶然与本质，描述统计与推断统计并重，在统计信息生成过程中防止数学游戏。统计分析报告和统计信息论文要采用通俗易懂、有骨有肉的数字语言，不能写成

“天书”。

传统的资料统计时代已经离去，代之而来的是信息资讯的新纪元。我们要充分利用电脑技术，加速推广通用软件，鼓励研制专用软件，有效地解决秩和比法的推广应用问题。作者主张极力开发统计信息产业，有这样一个设想，通过协作试编一个模拟数字试验软件，利用 RSR 进行分档并作相应的统计检验，既是一种智力游戏，又是一种科学实验方法的操练。

本书，是科研成果的体现，是应用的普及本，面向大众，面向未来。全书采用新的观念、结构、方法，以示例和评注的形式来编写，共选用二百几十个实例，均采用秩和比法做了处理。

秩和比法最大的特点是古典的参数统计与近代的非参数统计融合的最佳切入点，代表着统计信息方法未来的发展方向。

最后，我要衷心感谢中国预防医学科学院卫生信息中心的领导和同志们的大力支持与帮助，金水高教授在本书定稿期间虽屡屡出差国内外，但始终关照本书的编写；马林茂老师亲自审核稿件；马素香同志承担部分电脑文字处理；陶庄同志协助完成电脑制图和校对；中国协和医院疾病分类合作中心董景五、刘爱民老师热情接受有关咨询，支持出书不遗余力。我的学生金水高、廖建岗、梅林、段银康、姜垣在我出书遇到困难的情况下，伸出援助之手，从而最终实现了我出此书的夙愿。在此深表谢意！我的太太傅凌云女士为本书实例承担全部编秩工作；我女儿的好友郑捷女士为本书电脑制作和修订做出了贡献；女婿张维信先生在文稿设计和出书过程中给予了诸多技术帮助。对于这些，我深表敬意和谢意。

由于作者在认识上的局限性，势必有不恰当的地方甚至出现错误之处，尚望广大读者给予指正，在此表示衷心感谢！

作 者

2002 年 中秋节

目 录

第一章 秩和比法	1
第一节 从一个新的统计量——秩和比说起	1
一、秩和比的计算	1
二、“秩和比”多功能的体现	2
第二节 秩和比法	9
一、编秩	9
二、平均秩次法——RSR 的分布与切割	20
三、怎样做比较——综合评价及有关问题	23
四、怎样找关系——和谐分析及有关问题	37
五、秩和比法中的几项有关的技术问题	52
六、秩和比法是问卷调查的良好配套方法	56
七、利用秩和比法进行统计监控与测报的探讨	58
第三节 秩和比法与有关数量方法的比较研究	63
一、秩和比法与非参数统计的比较	64
二、秩和比法与多元分析的比较	78
三、秩和比法与模糊数学、灰色系统的比较	84
四、秩和比法与统筹方法的比较	89
五、秩和比法与时间数列分析的比较	95
第四节 秩和比法的功能	101
一、秩和比法的基本功能——统计研究与统计管理	101
二、统计信息资源的开发利用	101
第二章 卫生统计信息论的构想	107
第一节 疾病防治统计信息	107
一、国际疾病分类的推广应用	107
二、传染病和寄生虫病(AB)防治统计信息示例	108
三、恶性肿瘤(C)防治统计信息示例	120
四、良性肿瘤等类疾病(DEF GH)防治统计信息示例	125
五、循环系、呼吸系及消化系疾病(IJK)防治统计信息示例	135
六、皮肤和皮下组织疾病等(LMNOPQ)防治统计信息示例	154

七、症状等(RSTVWXYZ)防治统计信息示例	156
第二节 卫生管理统计信息	164
一、卫生资源统计信息	165
二、卫生服务统计信息	174
三、医教与医学科技管理统计信息	199
第三节 医学统计信息	202
一、临床医学统计信息	202
二、基础医学统计信息	216
三、预防医学与公共卫生统计信息	235
第四节 健康统计信息	276
一、人群及其结构统计信息示例	276
二、婚姻和生育统计信息示例	277
三、死亡和疾病统计信息示例	280
四、身体发育统计信息示例	283
五、寿命与生活质量统计信息示例	286
六、健康趋势与人口综合统计信息示例	288
第三章 秩和比法在认识缤纷世界上的应用	293
第一节 秩和比法在热门话题评说中的应用	293
一、环境保护	293
二、人口问题与计划生育工作	303
三、中国现代化进程的研究探索	314
第二节 秩和比法在工农业生产、交通、金融、商贸评价中的应用	328
一、工农业生产、交通	328
二、金融商贸	335
第三节 秩和比法在国民经济与社会生活评价中的应用	346
一、资源分布的考评	346
二、宏观经济与微观经济的分析	348
三、社会生活的评价	357
第四节 秩和比法在文教政法等中的应用	366
一、文教科技	366
二、卫生体育	374
三、政法民政	379

第一章

秩和比法

第一节 从一个新的统计量——秩和比说起

秩和 (rank sum), 是早就有的一个名词, 顾名思义是秩次总和的意思; 秩和比 (rank sum ratio, RSR) 却是一个新创的名词, 简单地说, 秩和比是秩和的一个特殊的平均数。

在我们看来, 秩和比是秩统计中一个内涵极为丰富的统计量, 其特征是 0—1, 表明不同计量单位多个指标的平均综合水平, 同时又是一个高度概括的综合指数, 或称复合信息的载体。秩和比是非参数统计通向参数统计的“接口”、“切入点”, 也可说秩和比是朴素秩和的发展与深化。

一、秩和比的计算

秩和比 (RSR) 的基本形态以下式表达:

$$RSR = \frac{\sum R}{m \cdot n}$$

式中 n 为分组数, m 为指标数, R 为秩次。

按行计与按列计秩和比以及加权秩和比的计算:

$$\text{按行计, } RSR_R = \frac{\sum_{\downarrow} R}{m \cdot n} \quad (\text{通常箭头不标出})$$

$$\text{按列计, } RSR_C = \frac{\sum_{\rightarrow} R}{m \cdot n} \quad (\text{通常箭头不标出})$$

$$\text{加权秩和比: } RSR_W = \frac{\sum WR}{n}$$

$$\text{最小 } RSR_R \text{ (或最小 } RSR_C) = \frac{1}{n} \text{ (或 } \frac{1}{m})$$

$$\text{最大 } RSR_R \text{ (或最小 } RSR_C) = 1$$

单指标 $RSR = \frac{R}{n}$ 。由于 n 的不同, RSR_R 的最小值也不一样。如: $n = 3$ 时,

$RSR_{min} = .3333$; $n = 50$ 时, $RSR_{max} = .02$ 。

一般说来, 行列 RSR 的计算有无错误可用如下方法来验证:

$\sum_1^n (RSR_R)$ 或 $\sum_1^m (RSR_C) =$ 一个整数或带有 0.5 的数, 如果 RSR 的计算不是依照基本形态, 则结果不受此约束。

RSR 的基本特征: 设定算得的 RSR 值越大越好, 也即 RSR 值永远为高优指标, 为此必须妥善解决编秩的问题。

RSR 的特点: RSR 值为无量纲, 综合能力强, 可代替一些专用综合指数, 可容纳一些专用统计量的信息, 如样本量 n , s 与 cv 等, 一些百分位数或流行病学指标 (如 OR、RR、PAR 等); 可显示微小变动; 异常值问题迎刃而解; 由于指标值采用秩代换, 会丧失一些信息, 且要求 $m \cdot n$ 矩阵中各元素不能缺少 (如个别缺, 可估) 是其局限性。此外, 在定水平方面也还是个薄弱环节; 统计测报中的应用也往往受到制约。

二、“秩和比”多功能的体现

(一) RSR 寓有极强的综合力

我们认为秩和比是一个通用的综合指数, 更是一个使用简便、效果良好的综合器。综合指数是用一个指标来概括多个指标的综合水平, 集中反映出各个侧面 (计量单位不同、类别功能各异) 的综合作用。秩和比还可称之为复合信息的载体, 寓有极强的综合力。

例 1-1 五所医院阑尾切口化脓率两年比较

表 1-1 五所医院阑尾切口化脓率两年比较 (R↓)

医院	1978 年 化脓率	1980 年 化脓率	化脓率下降 % $\frac{78-80}{78} \times 100\%$	$RSR = \frac{\sum R}{5 \times 3}$	$y = \sin^{-1} \sqrt{RSR}$	y 的 95% CL
本院	8.4 (1)	5.1 (3.5)	39.3 (5)	.6333	52.76	38.26-67.26
一院	5.0 (5)	3.2 (5)	36.0 (4)	.9333	75.03	60.53-89.53
二院	7.2 (3)	6.0 (2)	16.7 (2)	.4667	43.09	28.59-57.59
三院	8.1 (2)	6.5 (1)	19.8 (3)	.4000	39.23	24.73-53.73
四院	5.3 (4)	5.1 (3.5)	3.8 (1)	.5667	48.83	34.33-63.33
编秩准则	低优	低优	高优			

$n = 15$, $1.96Sy = 14.50$

评注: 我们在原有数据的基础上, 增列出化脓率下降百分比。增列原因一方面是强化两年化脓率的变化速度, 另一方面可增大 N (由原来的 $5 \times 2 = 10$ 个格子变为 $5 \times 3 = 15$ 个格子)。增列什么指标是个难点, 由分析者自行掌握, 一般说来, 这和业务与统计信息工作的熟悉程度有关。最后用 RSR 值的可信区间进行医院间的比较, 参见图 1-1。

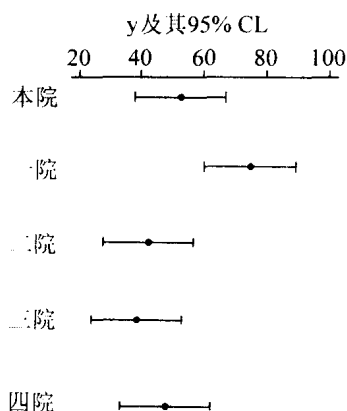


图 1-1 五所医院阑尾切口
化脓率的比较

分析表明：最好的是一院，最差的是三院（两者相差显著），本院与二院、四院居中（三者相差不显著）。1978 与 1980 两年一起考虑不仅可提高统计信息的稳定性，而且支持原来的看法，并有补充的新信息出现。

图示技术是秩和比法中常遇到的问题。它不仅在信息表达上有不可忽视的作用，很多时候，由于图示直观，可以帮助人们进行分析比较，利用 RSR 的（通过 $y = \sin^{-1} \sqrt{RSR}$ ）95% 可信区间作比较时常配以图示。本例五所医院两两比较，如无图的帮助，很难在较短时间内获得一个清晰的认识。

例 1-2 三个直辖市 1992 年健康状况与卫生资源的设置、利用的综合分析。

表 1-2-1 三个直辖市三类指标的综合分析 (R) (一)

指 标	健康 状 况 (A)			编秩准则
	北 京	上 海	天 津	
出 生 率	7.98 (2)	7.27 (3)	11.32 (1)	低优
死 亡 率	5.54 (3)	7.06 (1)	5.93 (2)	低优
自然增长率	2.44 (2)	.21 (3)	5.39 (1)	低优
婴儿死亡率	12.12 (1)	11.42 (2)	10.72 (3)	低优
新生儿死亡率	8.98 (1)	7.15 (3)	7.78 (2)	低优
孕产妇死亡率	3.00 (1)	2.85 (2)	1.21 (3)	低优
产妇死亡率	1.92 (2)	2.09 (1)	1.11 (3)	低优
e_0M	72.56 (1)	74.04 (3)	73.02 (2)	高优
e_0F	76.41 (2)	77.91 (3)	75.21 (1)	高优
$RSR_A = \frac{\sum R}{3 \times 9}$	.5556	.7778	.6667	$n = 3 \times 9 = 27$
y_A	48.19	61.88	54.74	$1.96Sy = 10.81$
y_A 的 95% CL	37.38-59.00	51.07-72.69	43.93-65.55	

表 1-2-2 三个直辖市三类指标的综合分析 (R) (二)

指 标	卫 生 资 源 的 设 置 (B)			编秩准则
	北 京	上 海	天 津	
平均每千人口 床位数	市区 7.31 (3)	市区 6.15 (2)	市区 5.75 (1)	高优
	县 3.39 (2)	县 4.43 (3)	县 1.52 (1)	
平均每千人口 医院床位数	市区 6.88 (3)	市区 5.61 (2)	市区 5.45 (1)	高优
	县 3.28 (2)	县 3.97 (3)	县 1.39 (1)	
平均每千人口 卫技人员数	市区 13.87 (3)	市区 11.43 (2)	市区 10.53 (1)	高优
	县 5.19 (2)	县 5.56 (3)	县 2.62 (1)	

续表

指 标	卫 生 资 源 的 设 置 (B)			
	北 京	上 海	天 津	编秩准则
平均每千人口 医生数	市区 6.27 (3) 县 2.62 (2)	市区 5.67 (2) 县 2.79 (3)	市区 4.84 (1) 县 1.45 (1)	高优
平均每千人口 护士数	市区 4.52 (3) 县 1.41 (2)	市区 3.54 (2) 县 1.73 (3)	市区 3.12 (1) 县 .58 (1)	高优
$RSR_B = \frac{\sum R}{3 \times 10}$	.8333	.8333	.3333	$n = 3 \times 10 = 30$
y_B	65.90	65.90	35.26	$1.96Sy = 10.25$
y_B 的 95% CL	55.65-76.15	55.65-76.15	25.01-45.51	

表 1-2-3 三个直辖市三类指标的综合分析 (R) (三)

指 标		卫 生 资 源 的 利 用 (C)			
		北 京	上 海	天 津	编秩准则
县及县以上 医院	使用率	86.8 (1)	91.2 (3)	87.6 (2)	高优
	周转次数	12.9 (3)	12.7 (1.5)	12.7 (1.5)	高优
	平均住院日	25.0 (2)	25.7 (1)	24.4 (3)	低优
卫生部门	使用率	90.4 (2)	93.3 (3)	90.3 (1)	高优
	周转次数	13.5 (3)	13.0 (1)	13.4 (2)	高优
	平均住院日	25.2 (2)	25.6 (1)	24.4 (3)	低优
工业及 其他部门	使用率	75.3 (2)	78.3 (3)	75.1 (1)	高优
	周转次数	11.2 (3)	10.6 (2)	9.5 (1)	高优
	平均住院日	24.9 (2)	26.5 (1)	23.9 (3)	低优
乡卫生院	使用率	59.7 (2)	76.6 (3)	49.5 (1)	高优
	周转次数	22.6 (1)	49.0 (3)	26.2 (2)	高优
	平均住院日	8.3 (1)	5.2 (3)	6.3 (2)	低优
$RSR_c = \frac{\sum R}{3 \times 12}$		.6667	.7083	.6250	$n = 3 \times 12 = 36$
y_c		54.74	57.31	52.24	$1.96Sy = 9.36$
y_c 的 95% CL		45.38-64.10	47.95-66.67	42.88-61.60	

表 1-2-4 三个直辖市三类指标的合计

指 标	三 类 指 标 的 合 计		
	北 京	上 海	天 津
$\sum R_{(ABC)}$	64	71.5	50.5
$RSR_{ABC} = \frac{\sum R}{3 \times 31}$	.6882	.7688	.5430
$y_{(ABC)}$	56.06	61.26	47.47
$y_{(ABC)}$ 的 95% CL	50.24-61.88	55.44-67.08	41.65-53.29

 $n = 93, 1.96Sy = 5.82$

分析表明：就三类指标合并计算，上海优于北京；北京优于天津；且均显著。在健康状况上，上海优于北京。在床人设置上，北京、上海并列榜首；天津最差。在床位利用上，三市相差均不显著。

评注：综合分析的目标是提取综合统计信息，在方法上没有一定的模式可循。综合指数 RSR 寓有极强的概括能力，因此，我们要学会如何构造 RSR 的思想与方法来解决各种实际问题。

(二) RSR 有时也表达为独立事件的概率

例 1-3 陕西秦巴山区残疾人综合考察背景的再分析

数据来源：颜虹等；陕西秦巴山区“痴呆、傻、聋、哑”综合科学考察流行病学，第二届全国医学人口学术研讨会交流，1991。

表 1-3-1 陕西秦巴山区残疾人不同背景下发生的概率及其组合 (一)

A	自然条件	RSR _A	B	社会发展	RSR _B	C	人民总收入	RSR _C	D	母亲受教育年限	RSR _D
A ₁	平川	.5000	B ₁	好	.5000	C ₁	<337	.1000	D ₁	0	.1667
A ₂	丘陵 低山	.3333	B ₂	中	.3333	C ₂	337-	.2000	D ₂	1-	.3333
A ₃	中高山	.1667	B ₃	差	.1667	C ₃	640-	.3000	D ₃	≥8	.5000
						C ₄	≥943	.4000			

表 1-3-2 陕西秦巴山区残疾人不同背景下发生的概率及其组合 (二)

AB 的组合		R	RSR = $\frac{R}{9}$
A ₁ B ₁	.5 × .5 = .25	平川，社会发展好	9 1.0000
A ₁ B ₂	.5 × .3333 = .1667	平川，社会发展中等	7.5 .8333
A ₁ B ₃	.5 × .1667 = .0834	平川，社会发展差	4.5 .5000
A ₂ B ₁	.3333 × .5 = .1667	丘陵低山，社会发展好	7.5 .8333
A ₂ B ₂	.3333 × .3333 = .1111	丘陵低山，社会发展中等	6 .6667
A ₂ B ₃	.3333 × .1667 = .0556	丘陵低山，社会发展差	2.5 .2778
A ₃ B ₁	.1667 × .5 = .0834	中高山，社会发展好	4.5 .5000
A ₃ B ₂	.1667 × .3333 = .0556	中高山，社会发展中等	2.5 .2778
A ₃ B ₃	.1667 × .1667 = .0278	中高山，社会发展差	1 .1111

评注：通过 RSR 的计算，扩大了对陕西秦巴山区残疾人不同背景的认识。在 $9 \times 12 = 108$ 种所有组合中：最低者—— $A_3B_3C_1D_1 = 0.009255$ ，最高者—— $A_1B_1C_4D_3 = 1.0000$ 。

本例与上例（例 1-2，例 1-3）表明，统计表的组织与通常的统计表不同，主项与宾项的设置存在着极大的灵活性，读者注意。

表 1-3-3 陕西秦巴山区残疾人不同背景下发生的概率及其组合 (三)

CD 的组合			R	$RSR = \frac{R}{12}$
$C_1 D_1$	$.1 \times .1667 = .0167$	收入差, 母文盲	1	.0833
$C_1 D_2$	$.1 \times .3333 = .0333$	收入差, 母文化中等	2.5	.2083
$C_1 D_3$	$.1 \times .5 = .05$	收入差, 母文化高	4.5	.3750
$C_2 D_1$	$.2 \times .1667 = .0333$	收入中等, 母文盲	2.5	.2083
$C_2 D_2$	$.2 \times .3333 = .0667$	收入中等, 母文化中等	6.5	.5417
$C_2 D_3$	$.2 \times .5 = .1000$	收入中等, 母文化高	8.5	.7083
$C_3 D_1$	$.3 \times .1667 = .05$	收入良, 母文盲	4.5	.3750
$C_3 D_2$	$.3 \times .3333 = .1$	收入良, 母文化中等	8.5	.7083
$C_3 D_3$	$.3 \times .5 = .15$	收入良, 母文化高	11	.9167
$C_4 D_1$	$.4 \times .1667 = .0667$	收入优, 母文盲	6.5	.5417
$C_4 D_2$	$.4 \times .3333 = .1333$	收入优, 母文化中等	10	.8333
$C_4 D_3$	$.4 \times .5 = .2$	收入优, 母文化高	12	1.0000

(三) RSR 寓有标化意

我们知道, 由于人口性别、年龄等构成的不同和分组死亡率的差异, 两地总死亡率的差异, 两地总死亡率的比较经常出现种种错综复杂的情况, 直接应用总死亡率进行比较显然是不合理的, 此时通常采用标化法处理, 以标化死亡率进行比较。值得注意的是秩和比也具有和标化相同的功能。

例 1-4 两县食管癌死亡率的比较

数据来源: 杨树勤主编: 医学百科全书医学统计分卷, 1985, P11。

表 1-4-1 两县食管癌死亡率的比较 (R): RSR 法的处理

年 龄 (岁)	甲 县			乙 县		
	人口数	死亡数 d	死亡率 p	人口数	死亡数 d	死亡率 p
0-	378977	2	.0000053 (1)	282762	1	.0000035 (2)
30-	63436	11	.000173 (1)	39443	4	.000101 (2)
40-	54910	55	.001002 (1)	40488	29	.000716 (2)
50-	41970	151	.003598 (1)	33309	99	.002972 (2)
60-	25060	163	.006504 (1)	23167	122	.005266 (2)
70-	10780	70	.006494 (2)	14548	98	.006736 (1)
计	575133	452	.000786	433717	353	.000814
$RSR = \frac{\sum R}{6 \times 2}$		.5833			.9167	
y		49.80			73.22	
y 的 95% CL		38.34-61.26			61.76-84.68	

$$n = 6 \times 4 = 24, \quad 1.96S_y = 11.46$$

以 RSR 表达 (通过 y) 的两县食管癌标化死亡率相差显著 (图示略)。

用直接法求甲、乙两县食管癌标化率 (P') 及其 95% 可信区间*, 结果如表 1-4-2 与表 1-4-3 所示。

表 1-4-2 甲县 p' 及其 95% CL 的计算

年 龄 (岁)	N^{Δ}	n	p	$1-p$	NP	$\frac{N^2 P (1-P)}{n}$
0-	661739	378977	.0000053	.9999947	3.51	6.123986
30-	102879	63436	.000173	.999827	17.80	28.859489
40-	95398	54910	.001002	.998998	95.59	165.904985
50-	75279	41970	.003598	.996402	270.85	484.065870
60-	48227	25060	.006504	.993496	313.67	599.716618
70-	25328	10780	.006494	.993506	164.48	383.942170
计	1008850	575133			865.90	1668.613118

$$p' = \frac{865.90}{1008850} = .0008583$$

$$Sp' = \frac{\sqrt{1668.613118}}{\sum N} = .0000405$$

$$1.96Sp' = .0000794$$

$$p' \text{ 的 } 95\% \text{ CL: } .0008583 \pm .0000794 = .000779-.000938$$

表 1-4-3 乙县 p' 及其 95% CL 的计算

年 龄 (岁)	n	p	$1-p$	NP	$\frac{N^2 P (1-P)}{n}$
0-	282762	.0000035	.9999965	2.32	5.420245
30-	39443	.000101	.999899	10.39	27.099485
40-	40488	.000716	.999284	68.30	160.825226
50-	33309	.002972	.997028	223.73	504.129658
60-	23167	.005206	.994794	251.07	519.933759
70-	14548	.006736	.993264	170.61	295.029387
计	433717			726.42	1512.437760

$$p' = \frac{726.42}{1008850} = .0007200$$

$$Sp' = \frac{\sqrt{1512.437760}}{\sum N} = .0000385$$

$$1.96Sp' = .0000756$$

$$p' \text{ 的 } 95\% \text{ CL: } .0007200 \pm .0000756 = .0006444-.0007956$$

分析表明: 两县食管癌标化死亡率相差显著, 甲县高于乙县, 与 RSR 法所得结果一样。

* 中国医学科学院卫生研究所等主编: 卫生统计学, 人民卫生出版社, 1982 年第一版第三次印刷, p. 46, 98

(四) RSR 寓有鉴别力

例 1-5 一则消息的启示 北京晚报 1995 年 10 月 12 日第 6 版以“运动处方”为题报道了如下消息,基础数据经过整理如表 1-5-1 所示。

表 1-5-1 冠心病患者一年后造影复查结果

组 别	病变消退%	病情不变%	进 展%
运 动 组	28	62	10
不运动组	6	49	45

此外,还报道了两组每周运动耗卡数,并提出日步行 3km 的措施,有观点有数据有办法,是一则值得信赖的消息。但两组相差在统计上无显著性,没有提及(作为一个记者,我们不应要求做统计检验)。

现试用 RSR 法进行统计处理,结果如表 1-5-2 所示。

表 1-5-2 冠心病患者一年后复查结果的 RSR 分析
(R↓): 病情不变按高优系列编秩

组 别	R _{病变消退}	R _{病情不变}	R _{有进展}	$RSR = \frac{\sum R}{2 \times 3}$	y	y 的 95% CL
运 动 组	高优 (2)	高优 (2)	低优 (2)	1.0000	90	76.77-103.23
	高优 (2)	偏高优 (1.75)	低优 (2)	.9583	78.22	64.99-91.45
	高优 (2)	稍高优 (1.625)	低优 (2)	.9375	75.52	62.29-88.75
不运动组	(1)	(1)	(1)	.5000	45	31.77-58.23
	(1)	(1.25)	(1)	.5417	47.39	34.16-60.62
	(1)	(1.375)	(1)	.5625	48.59	35.36-61.82

本例所有可能格子组合数 $N = \text{组数} \times \text{复查结果数} \times \text{可变组合数}$,

$$(2) \quad (3) \quad (3)$$

$$n = 2 \times 3 \times 3 = 18, 1.96Sy = 13.23。$$

如果编秩时旨在提高疗效标准上,将“病情不变%”按低优系列编秩,则结果如表 1-5-3 所示。

表 1-5-3 冠心病患者一年后复查结果的 RSR 分析 (R↓): 病情不变按低优系列编秩

组 别	R _{病变消退}	R _{病情不变}	R _{有进展}	$RSR = \frac{\sum R}{2 \times 3}$	y	y 的 95% CL
运 动 组	高优 (2)	稍高优 (1.375)	低优 (2)	.8958	71.17	57.94-84.40
	高优 (2)	偏低优 (1.25)	低优 (2)	.8750	69.30	56.07-82.53
	高优 (2)	低优 (1)	低优 (2)	.8333	65.90	52.67-79.13
不运动组	(1)	(1.625)	(1)	.6042	51.01	37.78-62.24
	(1)	(1.75)	(1)	.6250	52.24	39.01-65.47
	(1)	(2)	(1)	.6667	54.74	41.51-67.97

从表 1-5-3 可以看出:病情不变%如按低优系列编秩,出现运动组与不运动组相差

不显著的现象,因此“病情不变%”不能按低优系列编秩。

评注:本例应用的难点仍是N的确定问题。编秩时有两种情况出现:一类为绝对标准,另一类为相对标准。绝对标准是不变的,如病变消退%是高优指标,一年后复查进展%是低优指标,这是毋庸置疑的;相对标准是可变的,如病情不变%可编为低优系列中的偏低优或稍低优,但不能编为低优,这是基于控制病情恶化而设定的编秩准则。本例运动组均显著高于对应的不运动组,进一步支持了原记者的报道。

第二节 秩和比法

秩和比法是多单位多指标较佳的数量研究方法,有极大的柔韧性。本法始于编秩,在获得RSR值后,运用参数统计进行分析判断的一整套数量方法。本法的理论意义在于扩大了非参数统计的功能,并揭示了近代非参数统计与古典参数统计的结合点(切入点),使两者相互补充,相得益彰,为最终实现完全融合创造条件。

秩和比法也是一组全新的、广谱的统计信息方法,操作简便,使用效果明显,可满足多方面的需要,配合科技和进行管理的诸多需要。

秩和比法的内容很多,概括起来有编秩;计算统计量RSR;确定RSR的分布;RSR的分割、比较与评价;行列间和谐程度的测定;图示法及校正法等。秩和比法可用于描述,也可用于推断;可用于处理各种类型的资料(对行列众多的大资料尤具优势)。此法还可以与其他许多数量方法相互沟通、移植、嫁接。举例来说,和谐分析来自Kendall和谐系数,而计算和谐系数则一改过去利用秩和,直接用秩和比;又例如统计量RSR(或Topsis等),不仅可以镶在数理统计方法之中,也可镶在其他数量方法,如运筹学、灰色系统与模糊数学之中。

秩和比法本身所涉及的数学不多,适合于大多数统计信息人员的阅读与应用。应用时要充分体现专业要求,服从专业规律。不同学科、不同专业的理论对部门、专业统计有着最为重要的直接指导意义。应用时还要重视实践经验。例如在编秩问题上,是高优还是低优,取决于业务判断和经验的运用。如果读者已经了解一些数理统计知识,应用本法将有很大的方便,也会更加得到丰厚的信息享受。

一、编秩

(一) 编秩原则

充分体现专业要求;合理区分指标的属性,力求所编秩次无逻辑上的混乱,确保高优、低优、不分高低优及其种种组合的运用符合实际,认真考虑并体现研究目的,坚持用辩证的观点来编秩。

(二) 编秩的一般方法

编秩(ranking)也难,也不难。根据业务上的属性区分指标是高优还是低优,一般是不困难的。如人口平均预期寿命、受检率、合格率、治疗有效率、诊断符合率等均

可视为高优指标；发病率、死亡率、超标率、住院病死率、平均住院日等可视为低优指标。不分高优与低优也是常见的一种形式。运用高优、低优与不分高低的各种组合，在细微表达秩次的作用上也是值得提倡的。同时引入高（低）优与偏高（低）优如不能满足需要，最多再引入稍高（低）优即可。但有些场合区分也不那么简单，有时还要考察研究目的才能判定，如人口密度。如研究工业化进程，则编以高优；如研究环境保护，似编以低优为宜。总之，秩和比法用得与不好的关键是编秩是否恰如其分，符合实际。

这里秩次的层次划分是：高（低）优，偏高（低）优 = $1/2$ （高优或低优 + 不分），稍高（低）优 = $1/2$ （偏高优或偏低优 + 不分）。可以对称编，也可以不对称。一般说来，最多编列：低优、偏低优、稍低优、不分、稍高优、偏高优及高优等 7 个层次。

在 n 个指标中：高优指标的编法是以最大指标值编以最高的秩次 n ，次大的编以 $n-1$ ，……，次小的指标值编以 2，最小的编以 1；低优指标的编法与高优指标相反，最小的指标值编以最高的秩次 n ，次小的编以 $n-1$ ，次大的编以 2，最大的编以 1。如遇几个指标值相同，则均编以平均秩次。

例 1-6 80 年代 17 国人口年龄构成秩次的确定与分析

参见：第三章 秩和比法在认识缤纷世界上的应用，例 3-8。

例 1-7 中国 1990 年农村医疗机构服务范围的比较

表 1-7 中国 9 省(区)农村医疗机构服务范围的构成(R↓)

省(区)	服 务 半 径								RSR = $\frac{\sum R}{9 \times 7}$
	计	0-	1-	2-	3-	4-	5-	10-	
山 西	100.0	82.9(9)	7.5(3)	5.7(4.25)	1.0(5)	.5(6)	1.5(5.5)	.9(6)	.6151
内 蒙 古	100.0	46.2(1)	17.7(5)	6.2(4.5)	4.9(5)	5.6(4)	2.6(4.5)	16.9(1)	.3968
吉 林	100.0	54.1(2)	23.4(6)	8.9(5.875)	2.9(5)	3.6(4.25)	5.4(4)	1.8(4)	.4940
黑 龙 江	100.0	66.5(7)	21.7(5.5)	7.5(5)	1.7(5)	.6(5.75)	.7(6)	1.1(5)	.6230
江 苏	100.0	81.5(8)	13.6(4)	2.7(4)	.6(5)	1.2(5.5)	.1(7)	.2(7)	.6429
安 徽	100.0	61.1(4)	23.6(6.5)	8.9(5.875)	2.6(5)	1.5(5)	2.3(5)	.1(8.5)	.6329
广 东	100.0	66.0(6)	24.7(7)	6.6(4.75)	.5(5)	1.6(4.75)	.6(6.5)	.1(8.5)	.6746
云 南	100.0	57.5(3)	11.3(3.5)	8.6(5.375)	2.8(5)	3.3(4.5)	10.9(3)	6.4(2)	.4187
陕 西	100.0	63.7(5)	14.7(4.5)	8.6(5.375)	1.9(5)	1.4(5.25)	6.4(3.5)	3.6(3)	.5020
计	100.0	64.4	17.5	7.1	2.1	2.1	3.4	3.4	-
编秩准则		高优	偏高优	稍高优	不分	稍低优	偏低优	低优	

评注：本例在编秩时 7 个分组选用了 7 个层次，是最多的一个例子。

分析表明：9 省（区）农村医疗机构服务范围的水平偏低，并且相差不很大。如以百分计，在 40~70 分之间，最低为内蒙古 40 分，最高为广东 67.5 分。

（三）编秩时“标准”的引入

1. 按换算值 (X_i/X_s ，或 X_s/X_i) 编秩