

# 秩和检验法在动物生态学研究中的应用\*

宛新荣 钟文勤\*\*

(中国科学院动物研究所 农业虫鼠害综合治理研究国家重点实验室 北京 100080)

**摘要:** 秩和检验法是非参数统计法中的一种重要方法,是对成组数据  $t$ -检验的一个补充。在动物生态学领域的实验设计以及数据统计分析中有一定的实用价值,本文简要介绍这一方法的使用。

**关键词:** 秩和检验法;非参数统计;实验设计

**中图分类号:** Q141 **文献标识码:** A **文章编号:** 0250-3263(2003)01-63-04

## The Rank-sum Test and Its Application in Animal Ecology

WAN Xin-Rong ZHONG Wen-Qin

(Institute of Zoology, Chinese Academy of Sciences, Beijing 100080, China)

**Abstract:** The rank-sum test is an important non-parametric statistical method that can be used to analyze experimental data in animal ecology. The application of this test is briefly described.

**Key words:** Experimental design; Rank-sum test; Non-parametric statistical method

秩和检验法(rank-sum test)为非参数统计法中的一种重要方法<sup>[1-4]</sup>,是对成组数据  $t$ -检验法(pooled  $t$ -test)的一个补充。与成组  $t$ -检验法相比,秩和检验方法使用简便,且对象无须满足服从正态分布的特征,应用范围更广。本文简要介绍秩和检验的统计方法的适用范围以及应用步骤。

### 1 秩和检验法的实例

为了便于读者了解秩和检验法的适用范围,本节列举几个应用秩和检验的例子。

例1,采用陷阱法捕捉动物,捕获时记录时间和个体年龄(成体或幼体)。捕尽样地内所有个体后共捕获60只,按捕获时间先后顺序将个体进行排列(表1)。问:成体(adult A)是否比幼体(juvenile J)更易于为陷阱法所捕捉?

例2,比较某种药物处理对动物的胎仔数影响。药物处理组与对照组动物的胎仔数数据列于表2中。问:该药物是否对实验动物的胎仔数产生明显影响?

例3,采用无线电遥测技术对动物进行研究,比较带项圈组(无线发射器)与对照组(不带发射器)的动物寿命的差异。实验中,带发射器的动物为15个,对照组

为21个,动物的寿命以及秩值列于表3中。问:带发射器的动物寿命与对照组是否存在差异?

例4,采用两种饲料饲喂试验动物,饲养设定的时间后按照体重将动物分为A, B, C, D, 4个等级,试验结果列于表4中。问:2种饲料对动物体重生长有没有差异?

上述实例的共同点:两组不同处理的动物个体,具有一个可供比较的共性数值指标,按照这个指标可以对所有实验个体进行排序,并根据排序结果考察不同处理组间在该项指标上是否存在差异。一般涉及类似特征的实验设计多数可以采用秩和检验法进行统计。

### 2 秩和检验的统计步骤

现以表1为例介绍秩和检验法的统计流程,先按照

\* 中国科学院知识创新工程领域前沿项目,中国科学院生物特别支持资助项目(No. STZ-01-06);

\*\* 通讯作者;

第一作者介绍 宛新荣,男,33岁,副研究员;主要从事啮齿动物生态学研究工作。

收稿日期:2002-02-18,修回日期:2002-11-10

表 1 按捕获时间先后对个体赋秩

序号	年龄	秩	序号	年龄	秩	序号	年龄	秩
1	A	60	21	A	40	41	A	20
2	A	59	22	A	39	42	J	19
3	A	58	23	J	37	43	J	18
4	J	57	24	J	37	44	J	17
5	A	56	25	A	37	45	A	16
6	J	55	26	J	35	46	A	15
7	A	54	27	A	34	47	J	14
8	A	53	28	A	33	48	J	13
9	J	52	29	A	32	49	J	12
10	A	50.5	30	A	31	50	J	11
11	J	50.5	31	J	30	51	A	10
12	A	49	32	A	29	52	J	9
13	A	48	33	A	28	53	A	8
14	A	47	34	J	27	54	J	7
15	J	46	35	J	26	55	J	6
16	J	45	36	A	25	56	J	5
17	A	44	37	J	24	57	A	4
18	J	43	38	A	23	58	J	3
19	J	42	39	A	22	59	J	2
20	A	41	40	J	21	60	J	1

表 2 药物处理与对照组动物的胎仔数

序号	属性	胎仔数	秩	序号	属性	胎仔数	秩	序号	属性	胎仔数	秩
1	对照	12	18	7	处理	9	12	13	对照	7	5
2	对照	11	17	8	对照	9	12	14	处理	7	5
3	处理	10	15	9	对照	8	8.5	15	处理	7	5
4	对照	10	15	10	对照	8	8.5	16	处理	6	2.5
5	对照	10	15	11	对照	8	8.5	17	处理	6	2.5
6	处理	9	12	12	处理	8	8.5	18	处理	5	1

表 3 不同处理条件下动物的寿命

序号	类型	寿命	秩	序号	类型	寿命	秩	序号	类型	寿命	秩
1	对照	600	36	13	对照	372	24	25	对照	243	12
2	对照	568	35	14	对照	370	23	26	遥测	222	11
3	对照	532	34	15	遥测	365	22	27	对照	215	10
4	遥测	518	33	16	对照	334	21	28	遥测	212	9
5	对照	507	32	17	遥测	326	20	29	对照	206	8
6	对照	483	31	18	对照	315	19	30	遥测	204	7
7	遥测	480	30	19	对照	308	18	31	对照	195	6
8	对照	451	29	20	遥测	303	17	32	对照	184	5
9	对照	443	28	21	遥测	297	16	33	遥测	175	4
10	遥测	412	27	22	对照	292	15	34	对照	150	3
11	对照	406	26	23	对照	288	14	35	遥测	133	2
12	遥测	383	25	24	遥测	265	13	36	遥测	120	1

表 4 不同饲料对试验动物生长的影响

序号	组别	等级	秩	序号	组别	等级	秩	序号	组别	等级	秩
1	1	A	21	9	2	B	15	17	2	C	9.5
2	1	A	21	10	1	B	15	18	2	C	9.5
3	1	A	21	11	1	B	15	19	2	D	3.5
4	2	A	21	12	1	B	15	20	2	D	3.5
5	2	A	21	13	1	C	9.5	21	2	D	3.5
6	2	A	21	14	1	C	9.5	22	1	D	3.5
7	2	A	21	15	1	C	9.5	23	1	D	3.5
8	2	B	15	16	2	C	9.5	24	1	D	3.5

指定的指标(例 1 按照捕获时间的先后;例 2 按照胎仔数的大小;例 3 按照寿命的长短)对个体进行秩的排序(例如,例 1 中样本量为 60,因此最高秩为 60),每个个体的秩按照排列顺序依次降低,最末的个体秩为 1。如果某两个个体同时被捕获,则平分应得到的秩,例如表 1 中第 10 个个体与第 11 个个体同时捕获,平分所得到的秩,即  $(51 + 50)/2 = 50.5$ ;如果某三个个体在同一时间被捕获,则该三个个体也平分所应得到的秩,表 1 中的个体 23,24,25 就分享同样的秩,即  $(38 + 37 + 36)/3 = 37$ (类似的情形也出现在表 2、表 4 中)。按照这个方法,对所有的个体进行秩的赋值。

然后,按照检验对象的属性(年龄),将所有成体组个体的秩累加,即计算秩和。在秩和检验中,不同组别的样本量可以相等,如例 1 中成体与幼体个数均为 30;也可以不等,如例 2、例 3、例 4。秩和检验的统计量要以样本量较小的组别为基准,如果两个样本的量相等,则任意取一个,计算其秩和,进行统计分析。另外,在统计分析之前,要判断检验方式是单侧检验还是双侧检验,根据检验的方式建立零假设与备择假设。一般说来,检验差异显著与否为双侧检验,检验指定的一方是否高于或者低于另一方的情形即属于单侧检验。本文的例 1 属于单侧检验,显著水平  $\alpha = 0.025$ ;而例 2、例 3、例 4 均属于双侧检验,其显著水平  $\alpha = 0.05$ 。

此外,秩和检验的统计方法还跟样本量有关。一般地,如果实验数据中任意一方的样本量大于 10,则按照  $u$  检验法进行统计分析(如例 1、例 3、例 4);如果双方的样本量均小于或等于 10,则用查表法进行(例 2),下面依次介绍这两种方法。

**2.1 采用查表法进行统计分析** 如果双方的样本量均小于或等于 10,采用查表法。一般的生物统计类书籍均附有秩和检验表<sup>[1,3,4]</sup>。现以例 2 为例说明查表法的应用。

#### (1) 根据题意建立假设

零假设:药物处理组与对照组动物的胎仔数没有

差异;

备择假设:药物处理组与对照组动物的胎仔数存在差异。

#### (2) 选择样本量较少的组,计算其秩和

例 2 中药物处理组与对照组的样本量均为 9( $n_1, n_2$  分别表示两组的样本量,并规定  $n_1 \leq n_2$ ,此处  $n_1 = n_2 = 9$ ),因此可任意选取 1 组,不妨选择药物处理组,根据表 2 计算处理组的秩和:

$$T = 15 + 12 + 12 + 8.5 + 5 + 52.5 + 2.5 + 1 = 63.5$$

查秩和检验表,当  $n_1 = n_2 = 9$ ,显著水平  $\alpha = 0.05$  时:  $T_1 = 66, T_2 = 105$ 。

#### (3) 比较统计量与表值,给出结论

检验的标准是:如果  $T_1 < T < T_2$ ,接受零假设;反之,如果  $T < T_1$  或者  $T > T_2$ ,拒绝零假设,接受备择假设。

本例中  $T = 63.5 < 66 = T_1$ ,因此拒绝零假设,接受备择假设,即药物处理组与对照组胎仔数存在显著差异。

例 2 的数据也可以采用成组数据  $t$ -检验法处理,统计结果为:  $t = 2.28, df = 16, P = 0.036$ ,结论与秩和检验法结果一致。

**2.2 采用  $u$  检验法进行统计分析** 当任意一组样本量大于 10 时(例 1,例 3,例 4),秩和  $T$  近似服从正态分布,即可采用  $u$  检验法,统计量计算公式为:

$$u = \frac{T - \mu}{\sigma} = \frac{T - \frac{n_1 \cdot (n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}} \quad (1)$$

现以例 3 为例说明  $u$  检验法的使用步骤。

#### (1) 根据题意建立假设

零假设:遥测组与对照组动物寿命没有差异;

备择假设:遥测组与对照组动物寿命存在差异。

#### (2) 选择样本量较少的组,计算其秩和

例3中遥测组样本量为15( $n_1 = 15$ ),对照组为21( $n_2 = 21$ ),遥测组样本量较少,因此采用遥测组作为统计量分析。先根据表3计算遥测组的秩和: $T = 237$ 。

再根据公式(1)计算统计量  $u$ :

$$u = \frac{T - \frac{n_1 \cdot (n_1 + n_2 + 1)}{2}}{\sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}}} \\ = \frac{237 - \frac{15 \cdot (15 + 21 + 1)}{2}}{\sqrt{\frac{15 \cdot 21 \cdot (15 + 21 + 1)}{12}}} \\ = \frac{-40.5}{31.16} = -1.30$$

根据题意,本例为双侧检验,显著水平  $\alpha = 0.05$ ,查表得:  $u_{0.05} = -1.64$

由于  $u = -1.30 > u_{0.05} = -1.64$ ,因此接受零假设,即遥测组寿命与对照组没有显著差异。

如果采用成组数据  $t$ -检验法,结果为:  $t = 1.400$ ,  $df = 34$ ,  $P = 0.17$ ,结论是2组动物寿命不存在显著差异,这与秩和检验法的结论一致。

### 3 秩和检验法与成组 $t$ -检验法、符号检验法的比较

从上面的统计步骤可以看到,秩和检验是成组  $t$ -检验的一种简化。从应用范围来看,秩和检验法要比成组  $t$ -检验法范围更广,凡是能应用成组  $t$ -检验的场合,均可以使用秩和检验;反之,使用秩和检验的场合不一定都能使用成组  $t$ -检验法。秩和检验法也存在着省略一些量化的信息的缺陷。所以一般情况下,能用到成组  $t$ -检验的场合,通常不采用秩和检验法。但是,在一些特殊的场合下,有实验数据次序值而没有具体数值(如例1的情形),或者试验数据是按照等级来记录的(如例4),在这些场合下,成组  $t$ -检验法是无能为力的,只能采用秩和检验法。

除此之外,秩和检验法与成组  $t$ -检验法还有一个重要的差别:成组  $t$ -检验法要求实验数据符合正态分布,而秩和检验法对实验数据则无此项要求。不过,大多数生物学实验数据都符合正态分布<sup>[1,3]</sup>,因而这种细

微的差别在动物生态学实验数据的分析和处理过程中实际上意义不明显。

秩和检验法与符号检验法的区别类似于成组  $t$ -检验法与配对  $t$ -检验法之间的区别。符号检验法适用对象为对等实验的两个个体,而秩和检验法的数据则不具一一对应的特征,只有组间的差异而没有组内的差异。例如带发射器的个体作为一个属性,而对照组作为另外一个属性。相比之下,符号检验则是由一个个对等实验构成的,如动物在一次实验中的选食嗜好,多个选食实验的结果就构成符号检验法的数据来源<sup>[5]</sup>。此外,秩和检验法不要求实验组与对照组个数的相等,而在符号检验中,处理组与对照组数据个数必须相等。

### 4 小结

符号检验法是配对数据  $t$ -检验法的一个补充,秩和检验法也是成组  $t$ -检验法的一个重要补充<sup>[6]</sup>。了解这几种检验法的区别,便于在实验设计中灵活掌握和使用,这对于提高工作人员实验设计与统计分析能力是必要的。无论是配对数据  $t$ -检验法、成组数据  $t$ -检验法、符号检验法、秩和检验法,只是多种备择检验方法之一。在许多场合下,实验数据可以同时由多种统计方法来完成。至于如何选择最优的统计方法,还需要考察实验数据的其它一些属性(例如分布函数的数学特性等)以及检验的目的,这就要求研究人员对具体问题做具体分析,不能一概而论。

### 参 考 文 献

- [1] 杜荣寿. 生物统计学. 北京:高等教育出版社, 1987.
- [2] Underwood A J. Experiments in Ecology. London: Cambridge University Press, 1997. 1 ~ 212.
- [3] Jolicoeur P. Introduction to Biometry. New York: Plenum Publishers, 1999.
- [4] 孙山泽. 非参数统计讲义. 北京:北京大学出版社, 2000.
- [5] 宛新荣, 王梦军, 王广和等. 布氏田鼠对主要贮草种类的选食嗜好. 生态学杂志, 2002, 21(1): 65 ~ 66.
- [6] 宛新荣, 钟文勤. 符号检验法及其在动物生态学研究中的应用. 动物学杂志, 2002, 37(6): 56 ~ 59.