

·循证医学中的医学统计学问题·

中位数的置信区间和假设检验

李济宾, 张晋昕, 李伟栋

(中山大学公共卫生学院, 广州 510080)

[摘要] 对于非对称分布资料以及半定量资料、分布末端无确切数值或分布类型未知的资料, 秩和检验侧重于对其分布位置(即分布重心)的检验。为了与这类资料(特别是定量资料)统计描述的方法相衔接, 有时希望对组间的中位数进行假设检验, 此时秩和检验就不再适用。本文结合实例介绍基于样本中位数的置信区间估计和假设检验方法。

[关键词] 中位数; 置信区间; 假设检验

[中图分类号] R195.1 [文献标识码] A [文章编号] 1671-5144(2010)06-0365-04

Interval Confidence and Hypothesis Testing of Median

LI Ji-bin, ZHANG Jin-xin, LI Wei-dong

(School of Public Health, Sun Yat-sen University, Guangzhou 510080, China)

Abstract: The rank sum testing mainly focuses on the testing of distributional position (distributional center of gravity) for asymmetric distributional data, semi-quantitative data, without exact numerical values in the extremity of distribution, or unclearly distributional data. However, it is not suitable for rank sum testing when research purpose is to compare the differences of several sample medians between groups. This article mainly introduces the estimation of interval confidence and hypothesis testing based on sample medians according to living examples.

Key words: median; interval confidence; hypothesis testing

对于分布呈明显偏峰分布或分布末端无确切数值及分布类型不明的资料(如疾病的潜伏期、患者的住院天数、人体中重金属元素的含量、大气中某污染物的日平均浓度等), 一般用中位数描述资料的平均水平。对于这样的资料, 秩和检验是对样本所代表的总体的分布位置(即分布重心)进行检验, 如果要通过假设检验对比样本所代表总体中位数是否相等时, 秩和检验方法就不合适, 此时应选择基于中位数的统计推断^[1]。目前的教材不突出讲授中位数置信区间的估计和假设检验方法。本文结

合实例介绍中位数置信区间的估计和假设检验, 以满足实际应用的需要。

1 总体中位数置信区间的估计

对于某个来自非对称分布总体的随机样本, 样本中位数可以作为总体中位数的点估计值, 这种方法简单, 但是不能考虑抽样误差的影响, 很难评价准确度。中位数置信区间的估计可以解决这些问题。此外还可以根据两总体中位数置信区间是否重叠来推断两个总体中位数是否相等。

中位数置信区间的估计方法主要有以下几种。

1.1 样本中位数置信区间上下限百分位的专用公式^[2]

上限百分位=

$$\left[n + Z_{\alpha}^2 + \left(1 + Z_{\alpha} \sqrt{(n^2 - 1) / (n + Z_{\alpha}^2)} \right) \right] / \left[2(n + Z_{\alpha}^2) \right] \quad (1)$$

[基金项目] 中山大学教学改革研究课题重点资助项目(编号:中大教务[2009]108号-1-26)

[作者简介] 李济宾(1986-), 甘肃张掖人, 在读硕士研究生, 研究方向为统计学方法及其医学应用。

[通讯作者] 张晋昕, Tel: 020-87332453; E-mail: zhjinx@mail.sysu.edu.cn

下限百分位=

$$\left[n + Z_{\alpha}^2 - \left(1 + Z_{\alpha} \sqrt{(n^2 - 1) / (n + Z_{\alpha}^2)} \right) \right] / \left[2(n + Z_{\alpha}^2) \right] \quad (2)$$

(1)和(2)式中; n 表示样本含量; Z_{α} 由置信度决定, $Z_{0.05}=1.96, Z_{0.01}=2.58$ 。根据公式(1)和(2)算得两个百分位数,乘以样本含量 n 便得到总体中位数置信区间上下限所在的位次。然后把原始资料按从小到大排序,便可求得中位数的置信区间;也可利用百分位数计算公式计算中位数的置信区间。

1.2 样本含量 n 较大情况下置信区间的估计^[3]

(1)当样本量 n 较大时($n > 20$),计算总体中位数的 $1-\alpha$ 置信区间上下限的位次公式为:

$$S_1^* = \frac{n - Z_{\alpha} \sqrt{n}}{2}, S_2 = n - S_1 + 1 \quad (3)$$

其中: Z_{α} 为标准正态分布的双侧界值, $Z_{0.05}=1.96, Z_{0.01}=2.58$ 。 S_1, S_2 分别表示置信区间上下限的位次,这里 S_1 为 S_1^* 取整。将原始资料按从小到大排序,便可求得总体中位数的置信区间。这种计算方法适用于未分组的原始数据。

(2)当样本含量很大(如 $n > 100$),同时资料为分组的频数表资料时,用(3)式方法计算比较繁琐。这种情况下,可以采用正态近似法计算总体中位数的 $1-\alpha$ 置信区间,公式为:

$$M \pm Z_{\alpha} S_M, S_M = \frac{i \sqrt{n}}{2 f_M}, 95\% \text{置信区间} = M \pm 1.96 S_M \quad (4)$$

其中, S_M 为样本中位数的标准误, f_M 为中位数所在区段的频数。

例 1^[4] 从某工厂随机抽取 35 名工人,测得尿氟含量(mg/L),数据如下:

0.88	1.14	1.40	1.40	1.64	1.68	1.72
1.90	2.00	1.35	1.22	1.58	1.65	1.77
1.25	1.88	1.98	1.79	0.95	0.90	1.50
1.90	1.96	2.05	2.20	2.15	2.38	1.00
1.25	1.43	1.45	2.36	2.34	1.68	2.33

试计算该厂工人尿氟含量中位数的 95% 置信区间。

方法 1:根据中位数置信区间的百分位专用公式得:上限百分位=0.658 1,下限百分位=0.341 9。所以置信区间上限所在位次为 0.658 1×35=23.03,取整数为 23;置信区间下限所在位次为 0.341 9×35=11.97,取整数为 12。把原始资料按从小到大的顺序排列,便得总体中位数的 95% 置信区间为 (1.43, 1.88)mg/L。

方法 2: $S_1^* = \frac{35 - 1.96 \sqrt{35}}{2} = 11.7$,故置信区间上、

下限的位次分别为 $S_1=12, S_2=24$ 。对原始资料排序后,便得总体中位数的 95% 置信区间为 (1.43, 1.90)mg/L。

方法 3:将例 1 的数据编制频数表(表 1):

表 1 尿氟含量频数表

尿氟值(mg/L)	f_i	Σf_i	累计频率(%)
0.5~	3	3	8.57
1.0~	10	13	37.14
1.5~	14	27	77.14
2.0~	8	35	100.00
合计	35	-	-

$$M=1.68 \text{ mg/L}, S_M = \frac{i \sqrt{n}}{2 f_M} = \frac{0.5 \sqrt{35}}{2 \times 14} = 0.105 6 \text{ mg/L}$$

95% 置信区间: $M \pm 1.96 S_M = 1.68 \pm 1.96 \times 0.105 6$, 即为: (1.47, 1.89)mg/L。

对比三种计算方法得到的中位数 95% 置信区间,三者的计算结果有一定的出入,其中方法 1 和方法 2 计算的结果比较接近,方法 3 计算的结果与另外两种出入相对较大。因为方法 3 适用于大样本,本例样本含量 $n=35$,尚不是很大的样本,故本例不提倡使用方法 3。随着样本量的增加,三者的计算结果将趋于一样。对于样本量比较小的抽样(如 $n < 20$),可以用查表法(基于二项分布理论)推断总体中位数的 $1-\alpha$ 置信区间^[3]。此外,对于单个样本总体中位数 $1-\alpha$ 置信区间的估计,上下限位次的计算可以借助符号检验法、Wilcoxon 符号秩和检验法^[5]。

2 中位数的假设检验

2.1 两个大样本资料中位数的假设检验——Z 检验方法^[6]

大样本计算中位数标准误的公式为:

$$a = \frac{n}{2}, b = \frac{\sqrt{n}}{2}; c_1 = a + b, c_2 = a - b; S_M = \frac{d_1 - d_2}{2} \quad (5)$$

上式中 n 为样本含量,由 a 和 b 计算得到的 c_1 和 c_2 是两个百分位序,可以算出 c_1 位序的百分位数 d_1 和 c_2 位序的百分位数 d_2, S_M 为中位数的标准误。

基于两大样本的中位数 M_1 和 M_2 的 Z 检验公式为:

$$Z = \frac{M_1 - M_2}{\sqrt{S_{M_1}^2 + S_{M_2}^2}} \quad (6)$$

Z 为标准正态分布统计量。

例 2^[7] 某研究者欲比较 A、B 两菌种对小鼠

巨噬细胞吞噬功能的激活作用,将 43 只小鼠随机分为两组,用常规巨噬细胞吞噬功能的检测方法,

获得两组的吞噬率(%),数据见表 2。判断两组的吞噬率是否居于相同水平。

表 2 不同菌种对小鼠巨噬细胞的吞噬率

A 菌组(24 只)						B 菌组(19 只)				
64.0	45.0	66.0	64.0	70.0	70.0	51.0	70.0	62.0	61.0	89.0
70.0	56.0	76.0	66.0	63.0	71.0	51.0	70.0	68.0	61.0	91.0
95.0	56.0	60.0	66.0	67.0	74.0	54.0	70.0	68.0	71.0	93.0
64.0	57.0	93.0	66.0	73.0	74.0	54.0	59.0	61.0	87.0	

两个样本来自的总体为偏峰分布,用中位数描述两组的平均水平: $M_1=66.0\%$, $M_2=68.0\%$ 。以下是两个中位数的假设检验。

(1) 建立假设,确定检验水准

H_0 : 两个总体的中位数相等($M_1=M_2$)

H_1 : 两个总体的中位数不相等($M_1 \neq M_2$)

$\alpha=0.05$

(2) 计算检验统计量 Z 值

甲村:

$$a = \frac{24}{2} = 12.00, b = \frac{\sqrt{24}}{2} = 4.90$$

$$c_1 = a + b = 16.90, c_2 = a - b = 7.10$$

$$d_1 = 70.0\%, d_2 = 64.0\%$$

$$S_{M_1} = \frac{d_1 - d_2}{2} = 3.00\%$$

乙村:

$$a = \frac{19}{2} = 9.50, b = \frac{\sqrt{19}}{2} = 4.36$$

$$c_1 = a + b = 13.86, c_2 = a - b = 5.14$$

$$d_1 = 70.0\%, d_2 = 59.0\%$$

$$S_{M_2} = \frac{d_1 - d_2}{2} = 5.50\%$$

Z 检验:

$$Z = \frac{M_1 - M_2}{\sqrt{S_{M_1}^2 + S_{M_2}^2}} = \frac{68.0 - 66.0}{\sqrt{(3.00)^2 + (5.50)^2}} = 0.319$$

(3) 确定 P 值,作出统计推断

$Z=0.319 < 1.96$,按 $\alpha=0.05$ 水准, $P>0.05$,不拒绝 H_0 ,可以认为两组的吞噬率相同。

2.2 两个小样本资料中位数的假设检验——超越检验^[5]

把两个样本的观察值混合后从小到大排序,“非交叉重叠”部分观察值的个数称为超过数。所谓超越检验,就是根据超过数的多少推断两总体的分布位置(即变量值的倾向性)是否相同。这是一种快速的检验方法,适用于计量资料或顺序资料。

一样本观察值中小于另一样本最小值的个数称为左侧超过数。而大于另一样本最大值的个数称为右侧超过数。左、右侧超过数可能分属于两个样本,也可能属于一个样本。

对于两中心位置比较的超越检验的基本步骤为:①将两个样本观察值按由小到大排序,计算左、右侧超过数。②计算检验统计量 S :当左、右侧超过数分属于不同样本时, S 等于两侧超过数之和;属于同一样本时, S 等于两侧超过数之差(取绝对值)。③按两样本含量 n_1 和 n_2 ($n_1 < n_2$) 查表,若 $S \geq S_a$,则 $P < \alpha$;若 $S < S_a$,则 $P > \alpha$,按事先规定的检验水准作出统计推断。

例 3^[8] 某医师为研究血铁蛋白与肺炎的关系,随机抽查了肺炎患者和正常人若干名,并测得血铁蛋白($\mu\text{g/L}$)含量,问肺炎患者与正常成年人平均血铁蛋白含量是否居于相同水平?

肺炎患者	31					68					174			199	237	238			457	492	515	599		
正常人		34	43	43	44	47	47	47	52	54	68	95	132	172			177						277	294

(1) 建立假设,确定检验水准

H_0 : 肺炎患者和正常人血铁蛋白含量的中位数相同($M_1=M_2$)

H_1 : 肺炎患者和正常人血铁蛋白含量的中位数不同($M_1 \neq M_2$)

$\alpha=0.05$

(2) 计算检验统计量 S ,确定 P 值,作出统计推断
将两组观察值按从小到大排序(如上),左侧超过数为 1,右侧超过数为 4,属于同一个样本,故 $S=4-1=3$ 。 $N_1=10$, $n_2=16$,查“两中心位置差别检验用 S

界值表”得,双侧检验 $S_{0.05}=8, S < S_{0.05}$, 故 $P > 0.05$, 按 $\alpha=0.05$ 的检验水准, 不拒绝 H_0 , 可以认为肺炎患者与正常人血铁蛋白含量的平均水平相同。

当 n_1 或 n_2 超过“两中心位置差别检验用 S 界值表”范围的时候, 可按公式直接计算单侧 P 值 (取大于等于 S 的累计概率), 取双侧 P 值时应乘以 2。

$$p = \frac{n_1}{n_1 + n_2} \quad (7)$$

$$\text{当 } p \neq \frac{1}{2} \text{ 时, } P = \frac{p^2 - p + 1}{2p - 1} \left(\frac{p^{S+1}}{1+p} - \frac{(1-p)^{S+1}}{2-p} \right) \quad (8)$$

$$\text{当 } p = \frac{1}{2} \text{ 时, } P = \left(\frac{S}{2} + \frac{1}{3} \right) \frac{1}{2^S} \quad (9)$$

以例 3 计算, $p = \frac{10}{10+16} = 0.3846$, 单侧 $P = 0.241$, $P > 0.05$, 结论相同。

2.3 中位数比较的 χ^2 检验法^[9]

基本思想: 如果各组样本数据对应的总体分布相同, 则各组资料合并后计算所得的中位数的上下每组样本应各有一半数目。基于这样的假设, 可以检验中位数上下各组观察值数目的差异是否有统计学意义, 从而作出统计推断。

计算步骤: ①将各样本资料混合后按由小到大排序; ②求混合后数据的中位数 (M_d); ③对每组样本分别计算超过混合资料的中位数的数目, 记为 m_1, m_2, \dots, m_p ; ④用 m_1, m_2, \dots, m_p 和 n_1, n_2, \dots, n_p 列出 $2 \times C$ 表格 (表 3); ⑤用 χ^2 检验法或确切概率法, 对 $2 \times C$ 表进行检验。

以例 2 中的数据为例, 得表 3 所示中位数检验表。

表 3 中位数检验用 $2 \times C$ 表

样本	A	B	……	P	合计
取值 $> M_d$	m_1	m_2	……	m_p	$n_1 + n_2 + \dots + n_p - c$
取值 $\leq M_d$	$n_1 - m_1$	$n_2 - m_2$	……	$n_p - m_p$	c
合计	n_1	n_2	……	n_p	$n_1 + n_2 + \dots + n_p$

(1) 建立假设, 确定检验水准

H_0 : 两组吞噬率的中位数相同 ($M_A = M_B$)

H_1 : 两组吞噬率的中位数不同 ($M_A \neq M_B$)

$\alpha = 0.05$

(2) 计算检验统计量, 确定 P 值, 作出统计推断

两组数据混合后的中位数为 66.0%。得表 4 所示 2×2 表。

表 4 两组吞噬率的分布

样本	A 菌组	B 菌组	合计
取值 $> 66.0\%$	11(11.7)	10(9.3)	21
取值 $\leq 66.0\%$	13(12.3)	9(9.7)	22
合计	24	19	43

由 SPSS 计算得 $\chi^2 = 0.196, v = 1, P = 0.658 > 0.05$, 按 $\alpha = 0.05$ 水准, 不拒绝 H_0 。所以, 可以认为两组的吞噬率的中位数相同。而以例 3 中的数据, 用中位数的 χ^2 检验法, $\chi^2 = 5.850, P = 0.016 < 0.05$, 按 $\alpha = 0.05$ 水准, 拒绝 H_0 , 接受 H_1 。比较发现, 对于大样本的资料, χ^2 检验法与 Z 检验的结果基本一致, 而对于小样本的资料, χ^2 检验法和小样本中位数超越检验的结论矛盾。可见, 中位数的 χ^2 检验法对样本量也有一定的要求, 一般大样本情况下结论比较可靠。此外, 对于多组样本中位数的比较, Z 检验和超越检验将不再合适, 此时只能选用 χ^2 检验法。

中位数检验方法与秩和检验方法均属于非参数检验的范畴, 但是它们检验的侧重点不一样。中位数检验方法直接比较两组的中位数, 而秩和检验是检验样本所代表的总体分布重心是否相同。本文针对基于中位数的置信区间的计算和假设检验, 旨在让大家辨清秩和检验与中位数检验的异同, 以供广大医学和卫生统计工作者参考之用。在实际工作中应根据研究的目的, 采取相应的检验方法。

[参 考 文 献]

- [1] 王学生, 王革. 中位数检验法在偏态分布资料中应用[J]. 中国公共卫生, 2001, 17(5): 462-463.
- [2] 申杰, 汤小京. 总体中位数的估计及其应注意的问题[J]. 中国卫生统计, 1994, 3: 22.
- [3] 黄正南. 中位数的统计推断[J]. 湖南医学, 1988, 11(5): 43-45.
- [4] 师明中, 杨连胜. 中位数的符号秩和检验法[J]. 数理医药学杂志, 1997, 10(2): 146-147.
- [5] 程琮, 程玮, 范华. 单样本中位数置信区间估计[J]. 中国卫生统计, 2002, 19(4): 243-245.
- [6] 郭祖超. 医用数理统计方法[M]. 北京: 人民卫生出版社, 1988: 114-115.
- [7] 方积乾. 卫生统计学[M]. 北京: 人民卫生出版社, 2008: 193-194.
- [8] 方积乾. 生物医学研究的统计方法[M]. 北京: 人民卫生出版社, 2007: 114.
- [9] 成军, 孙关忠, 李早荣. 中位数检验方法在临床实验室中的应用与评价[J]. 陕西医学检验, 2000, 15(2): 6-8.

[收稿日期] 2010-11-05