

94.15(3)
167-169

介绍正态性检验简捷法

邓习明 周应胜

(湖北省广水市卫生防疫站, 432700)

R195.1

摘要 本文用矩法和简捷法对 150 名 7 岁男孩的身高资料进行了正态性检验对比, 其 S_k 值为 -0.235, 近似于 0, K_u 值为 0.249, 也极接近 0.263, 证明简捷法与较精确的矩法检验结果相同, 由于简捷法计算简便, 而且实用、可靠, 可广泛应用于儿少卫生研究中。

关键词 矩法 简捷法 正态性检验

在儿少卫生科研工作中, 常常遇到正态性检验, 采用的方法有正态概率纸目测法、D 检验法和矩法^[1], 但它们有的要一定条件, 有的计算复杂, 基层较难应用, 本文通过一组资料采用矩法和简捷法统计, 并对其性能与应用范围进行了探讨。

1 原理与方法

1.1 矩法: 亦称峰度、偏度检验法和动差法, 是用数学上矩的原理来检验峰度和偏度的一种方法。计算公式:

$$g_1(\text{偏度系数}) = \frac{n\sum fx^3 - 3\sum fx \cdot \sum fx^2 + 2(\sum fx)^3/n}{(n-1)(n-2)\{\sum fx^2 - (\sum fx)^2/n\}/(n-1)}^{3/2} \quad (1)$$

$g_1=0$ 时曲线对称, $g_1>0$ 为正偏态, $g_1<0$ 为负偏态, 其标准误称为偏度标准误。

$$\sigma_{g_1}(\text{偏度标准误}) = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \quad (2)$$

$$g_2(\text{峰度系数}) = \frac{(n+1)[n\sum fx^4 - 4\sum fx \cdot \sum fx^3 + 6(\sum fx)^2 \cdot \sum fx^2/n - 3(\sum fx)^4/n^2]}{(n-1)(n-2)(n-3)\{\sum fx^2 - (\sum fx)^2/n\}/(n-1)}^2 - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (3)$$

$g_2=0$ 时曲线为正态峰, $g_2>0$ 为尖峭峰, $g_2<0$ 为平阔峰, g_2 的标准误称为峰度标准误。

$$\sigma_{g_2}(\text{峰度标准误}) = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}} \quad (4)$$

因 g_1 和 g_2 的抽样分布近似正态分布, 可用 u 检验推断分布的正态性。

$$u_{g_1}(\text{偏度 } u \text{ 检验}) = g_1/\sigma_{g_1} \quad (5)$$

$$u_{g_2}(\text{峰度 } u \text{ 检验}) = g_2/\sigma_{g_2} \quad (6)$$

根据 u_{g_1} 和 u_{g_2} , 查 u 界值表 (t 界值表, $r=\infty$ 时), 得 P 值, 按 P 值大小作出推断结论。

2 常用的参考值范围与其对应的 u 值 (见表 1)表 1 单侧检验和双侧检验 u 值表

α	百分范围 ($1-\alpha$)%	u 值	
		单侧	双侧
0.10	90	1.2816	1.6449
0.05	95	1.6449	1.9600
0.01	99	2.3263	2.5758

1.2 计算 S_k 和 K_u 值的简捷法:^[2,3,4]

1.2.1 基本原理: 理论上正态分布的均数与中位数相等, 两者之差为 0, 若为近似正态分布, 则其差近似于 0。正态分布的陡坡度为 0.263, 若计算的陡坡度接近于 0.263, 若计算的陡坡度接近于 0.263, 则该分布近似于正态分布。整个计算过程只需使用百分位数公式。

计算公式:

$$S_k(\text{偏态程度}) = \frac{P_{95} + P_{10}}{2} - P_{50} \quad (7)$$

$$K_u(\text{陡坡度}) = \frac{Q}{P_{90} - P_{10}} = \frac{(Q_9 - Q_1)/2}{P_{90} - P_{10}} = \frac{(P_{75} - P_{15})/2}{P_{90} - P_{10}} \quad (8)$$

2 实例

2.1 例 1: 随机抽取的 150 名 7 岁健康男孩的身高的频数分布见表 2 第 (2) 栏, 试用矩法和简捷法作正态性检验。

2.1.1 矩法: H_0 : 总体服从正态分布 (即总体偏度系数 $r_1=0$, 且总体峰度系数 $r_2=0$); H_1 : 总体非正态分布 (即 $r_1 \neq 0$ 或/与 $r_2 \neq 0$); $\alpha=0.1$ 。根据统计基本知识求得的 $\sum fx$, $\sum fx^2$, $\sum fx^3$, $\sum fx^4$ 见表 2 第 (4)~(7) 栏。将表 2 中合计栏内各有关数字代入上述公式 (1)~(4) 得:

$$g_1 = -0.1331, \sigma_{g_1} = 0.1980, g_2 = 0.2564, \sigma_{g_2} = 0.3936$$

据公式 (5) 得:

$$u_{g_1} = -0.1331/0.1980 = -0.6722$$

查 u 界值表, 得 $P>0.1$ 。

据公式 (6) 得:

$$u_{g_2} = 0.2564/0.3936 = 0.6514$$

查 u 界值表, 得 $P>0.1$ 。

结论: 经偏度检验和峰度检验, 按 $\alpha=0.1$ 水准不拒绝 H_0 , 可以认为这 150 名 7 岁男孩的身高服从正态分布。

表 2 150 名男孩身高分布用矩法作正态性检验计算表

身高(cm)	f	x	fx	fx ²	fx ³	fx ⁴
(1)	(2)	(3)	(4)	(5)	(6)	(7)
106~	1	-7	-7	49	-343	2401
108~	2	-6	-12	72	-432	2592
110~	3	-5	-15	75	-375	1875
112~	6	-4	-24	96	-384	1536
114~	8	-3	-24	72	-216	648
116~	17	-2	-34	68	-136	272
118~	21	-1	-21	21	-21	21
120~	26	0	0	0	0	0
122~	25	1	25	25	25	25
124~	19	2	38	76	152	304
126~	10	3	30	90	270	810
128~	7	4	28	112	448	1792
130~	3	5	15	75	375	1875
132~	1	6	6	36	216	1296
134~136	1	7	7	49	343	2401
合计	150	-	12	916	-78	17848
	(n)		(Σfx)	(Σfx ²)	(Σfx ³)	(Σfx ⁴)

2.1.2 简捷法:

根据资料列频数分布表(见表 3),根据表 3,求得 $P_{10}=114.75$, $P_{25}=118.05$, $P_{50}=121.31$, $P_{75}=124.37$, $P_{90}=127.40$,将这些数据代入公式(7)和公式(8)得:

$$S_k = -0.235, K_u = 0.249$$

结论:因为 S_k 近似于 0, K_u 值近似于 0.263(仅相差 0.017),所以可以认为该样本属正态分布或近似正态分布。

2.2 例 2:1984 年我们调查了 180 名 18 岁健康男子血铅含量($\mu\text{g}/100\text{g}$),其频数分布见表 4,因要计算 95% 上限,试用简捷法作正态性检验。

表 3 150 名男孩身高的频数分布

身高(cm)	频数(f)	累计频数	累计频率(%)
106~	1	1	0.67
108~	2	3	2.00
110~	3	6	4.00
112~	6	12	8.00
114~	8	20	13.33
116~	17	37	24.67
118~	21	58	38.67
120~	26	84	56.00
122~	25	109	72.67
124~	19	128	85.33
126~	10	138	92.00
128~	7	145	96.67
130~	3	148	98.67
132~	1	149	99.33
134~136	1	150	100.00

表 4

180 名 18 岁男子的血铅含量的频数分布

血铅含量 ($\mu\text{g}/100\text{g}$)	2~	5~	8~	11~	14~	17~	20~	23~	26~	29~	32~	35~	38~	41~	44~	47~50	合计
频数	3	29	38	28	14	14	9	8	3	1	1	0	1	0	0	1	180

从表 4 可看出该资料不成正态分布。现先对原始数据进行对数变换,再用简捷法进行正态性检验。

进行对数变换后列出的《血铅测定值转换对数的频数计算表》见表 5。

根据百分位数公式,得

$$P_{10}=0.617, P_{25}=0.811, P_{50}=1.009, P_{75}=1.189, P_{90}=1.343.$$

根据公式(7)和公式(8),得:

$$S_k = -0.029$$

$$K_u = -0.260$$

结论:本例的 S_k 比例 1 更接近于 0, K_u 也比例 1 更接近于 0.263,所以可以判定该资料经对数转换后成正态分布或近似正态分布。

3 讨论

不少统计方法仅适用于正态公布或近似正态分布。本文应用简捷法和较精确的矩法对一实例进行了正态性检验对比,结论完全相同。尽管简捷法的个别地方也有待完善。但作为一种卫生统计方法,计算 S_k 和 K_u 值不仅简便、实用,而且十分可靠。

在儿少卫生中,通过正态性检验,可以确定同年龄、同性别儿童的身高、体重、血红蛋白含量和红细胞数

表 5 血铅测定值转换对数的频数计算表

对数组段	真数组段	频数(f)	累计频数	累计频率(%)
(1)	(2)=lg ⁻¹ (1)	(3)	(4)	(5)
0.3~	2.0~	1	1	0.56
0.4~	2.5~	5	6	3.33
0.5~	3.2~	9	15	8.33
0.6~	4.0~	18	33	18.33
0.7~	5.0~	10	43	23.89
0.8~	6.3~	19	62	34.44
0.9~	8.0~	26	88	48.89
1.0~	10.0~	23	111	61.67
1.1~	12.6~	27	138	76.67
1.2~	15.8~	18	156	86.67
1.3~	20.0~	14	170	94.44
1.4~	25.1~	7	177	98.33
1.5~	31.6~	2	179	99.44
1.6~1.7	39.8~50.1	1	180	100.00
合计	-	180(n)	-	-

胞数等形态和生理指标是否为正态分布。若资料是

正态或近似正态分布,就可以计算均数 \bar{X} ,标准差 S ,确定正常值范围 $\bar{X} \pm uS$ 和作 u 检验。本文例 1 的 $\bar{X} = 121.16\text{cm}$, $S = 4.96\text{cm}$,95%的可信限范围为 $121.16 \pm 1.96 \times 4.96\text{cm}$ 即 $111.45 \sim 130.87\text{cm}$ 。在实践中,有的资料可通过对数变换(如学生视力资料)、平方根变换(如学习环境的细菌数)或百分数的平方根反正弦变换(如某些多发病的患病率、白细胞分类计数),然后再用简捷法作正态性检验。

本简捷检验法还可用于基础医学、临床医学、流

行病学等各个医学领域。

参考文献

- [1] 杨树勤主编:卫生统计学,第二版,北京:人民卫生出版社,1988,56~60.
- [2] 王圣基:简捷统计方法讲座,1978
- [3] 王翔朴主编,卫生学,第三版:北京:人民卫生出版社,1990,252~254.
- [4] 张选群,毛宗福:临床科研分析,武测出版社,1992:39.

海口市中小学视力低下状况分析

杨来益

(海南省海口市中小学卫生保健所,570001)

1991 年结合国家学生体质调研工作,对本市 3000 多名中小學生进行了视力检查,现报告如下:

1 对象与方法

选取监测点校(3 所中学和 2 所小学)7~18 岁中小學生为对象,从中随机抽取城市汉族学生各年龄组男、女各 105 名。

检查方法与诊断标准均按国家有关规定进行。检测人员固定并经过统一培训,预测合格。检测时间为 1991 年 4 月 8 日至 5 月 8 日。

2 结果与分析

2.1 视力低下情况:在 2520 名学生中,检出视力低下眼 1834 只,视力低下率为 36.39%,其中男生为 28.73%,女生为 44.05%。女生显著高于男生。

各年龄组男、女学生视力低下率均随着年龄的增加而增加,女生均高于男生(秩和检验结果, $P < 0.01$)。

2.2 各年龄组学生视力低下率增长情况:7~18 岁视力低下率由 14.52%增加到 68.81%,12 年共增加 54.29 个百分点,平均每年增长 4.52 个百分点。12 岁、15 岁和 18 岁 3 个年龄组从数值上看增长率较

大,可能与这些接近毕业的学生,学习负担重、用眼过度以及体育活动减少等因素有关。

2.3 学生平均视力情况:检查 5040 只眼,视力平均值 4.885,其中,男生视力平均值 4.943,女生视力平均值 4.826,男生好于女生,差异非常显著($t = 12.84$, $P < 0.01$)。

2.4 视力低下程度:轻度视力低下占 17.12%,中度占 33.53%,重度占 49.35%。随着学习阶段的增加,视力低下构成比,轻、中度逐渐降低,而重度则快速升高。

2.5 屈光性近视眼与优等视力情况:

在视力低下眼中,屈光性近视占 93.68%,远视占 2.27%,其他占 4.05%。在 5040 只眼中,优等视力眼(视力 5.2 及 5.3 者称之)797 只,占 15.81%,在小学阶段,优等视力率右眼多于左眼,视力低下率右眼少于左眼,初中阶段上述关系逐渐改变,到高中阶段则相反,优等视力率右眼少于左眼,视力低下率右眼多于左眼,可能与学生中绝大部分为右斜,致使右眼近视工作比左眼多有关。

(1991-12-31 收稿)

小学生流行性腮腺炎发病报告

拓建国

(新疆兵团农二师卫生防疫站,841000)

1990 年 11 月 20 日~12 月 24 日,我市华山小学发生流行性腮腺炎 17 例,并发脑膜脑炎 1 例。

该校有 602 名学生,15 个班级。首例病人为男性,7 岁,有发热、头痛、咽痛、食欲不振、恶心,2 天后腮腺肿大,诊断为流行性腮腺炎。截止 12 月底共发

生 17 例,发病率为 2.8%(17/602)。均属 7 岁年龄组的一年级四班学生。发病占该班学生的 43.5%(17/39),男:7 例,占 33.3%(7/21),女:10 例,占女生的 55.5%(10/18)。并发脑膜脑炎 1 例,并发率为 5.8%。经积极治疗,无一例死亡。