

· 计算机应用 ·

用矩法进行正态性检验及其在统计软件中的实现

中国医学科学院基础医学研究所 北京协和医学院基础学院统计学教研室(100005) 徐 涛 韩少梅

有些统计方法只适用于正态分布资料或近似正态分布的资料,如用 t 检验和方差分析进行两个或多个样本均数的比较,每个样本都要满足正态性要求,进行直线回归要求应变量 y 来自正态总体,因此在进行资料的分析之前,首先要探讨样本资料是否来自正态总体,即正态性检验。但是在分析的过程中,经常会遇到这样的问题,同一资料用几种方法进行正态性检验,检验结果不一致,正态性检验结果和直方分布图、正态概率图不相符合,甚至出现完全相反的结果,选用哪种方法进行正态性检验更合适呢?

本文简要地介绍了统计软件常用的正态性检验方法存在的问题,详细描述了用矩法进行正态性检验的基本原理、步骤以及如何用常用的统计软件 SPSS 和 SAS 来实现该检验。

统计软件常用的正态性检验方法存在的问题

在 SPSS 和 SAS 等统计分析软件中,通常用统计描述模块中的 Shapiro-Wilk 检验、经过 Lilliefors 显著水平修正的 Kolmogorov-Smirnov 检验和非参数检验模块中的单一样本 Kolmogorov-Smirnov 检验进行正态性判定。但是这几种检验方法存在以下几方面的问题。

(1)在实际应用中常出现检验结果与直方图、正态性概率图不一致,甚至几种假设检验方法结果完全不同的情况。

(2)Shapiro-Wilk 检验(W 检验)和经过 Lilliefors 显著水平修正的 Kolmogorov-Smirnov 检验(D 检验)是用一个综合指标(顺序统计量 W 或 D)来判定资料的正态性^[1]。由于两种方法都是用一个指标反映资料的正态性,所以当资料的正态峰和对称性两个特征有一个不满足正态性要求时,两种方法出现假阴性错

误的机率均较大;而且两种方法的检验统计量都是进行大小排序后得到,所以易受异常值的影响。

(3)Kolmogorov-Smirnov 单一样本检验是根据实际的累计频数分布和理论的累计频数分布的最大差异来检验资料的正态性,可对正态分布进行拟合优度检验。但它并非检验正态性的专用方法,因此它的检验效率是最低的,最容易受样本量和异常值等因素的影响^[2]。

为此,可以用矩法进行正态性检验。由于矩法是基于对称性和正态峰这两个基本特征来检验资料的正态性,所以它的检验结果是最接近直方图、正态概率图的,是检验资料正态性最恰当的方法。

矩法基本原理及分析步骤

矩法,亦称动差法、偏度峰度检验,是应用随机变量 X 的标准化变量的三阶矩和四阶矩作为偏度和峰度,对资料的偏度和峰度各用一个指标(偏度系数和峰度系数)来判定资料的正态性,因此用矩法进行正态性检验包括偏度检验和峰度检验两部分,只有偏度检验和峰度检验均得到 $P > 0.05$ 的结论时,才能认定该组资料符合正态分布。分别以 g_1 和 g_2 代表样本偏度系数和峰度系数,偏度系数等于 0 为对称,大于 0 为正偏态,小于 0 为负偏态;峰度系数等于 0 为正态峰,大于 0 为尖峭峰,小于 0 为平阔峰。

在实际应用中,因为抽样误差的缘故,偏度系数和峰度系数不会刚好等于 0,因此需要通过假设检验对偏度系数和峰度系数是否为 0 进行检验。该检验是基于 U 检验的原理,利用公式 $U_1 = g_1/\sigma_1$ 和 $U_2 = g_2/\sigma_2$ 得到 U 值,并与标准正态分布界值进行比较以分析资料的偏度和峰度。 σ_1 和 σ_2 为偏度系数和峰度系数的标准误。 g_1 、 g_2 、 σ_1 和 σ_2 的计算公式^[1,3-5]如下:

$$g_1 = \frac{n \sum fX^3 - 3 \sum fX \sum fX^2 + 2(\sum fX)^3/n}{(n-1)(n-2)\{[\sum fX^2 - (\sum fX)^2/n]/(n-1)\}^{3/2}} \quad (1)$$

$$g_2 = \frac{(n+1)[n \sum fX^4 - 4 \sum fX \sum fX^3 + 6(\sum fX^2)^2 \sum fX^2/n - 3(\sum fX)^4/n^2]}{(n-1)(n-2)(n-3)\{[\sum fX^2 - (\sum fX)^2/n]/(n-1)\}^2} - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (2)$$

$$\sigma_1 = \sqrt{\frac{6n(n-1)}{(n-2)(n+1)(n+3)}} \tag{3}$$

$$U_1 = g_1/\sigma_1 \tag{4}$$

$$\sigma_2 = \sqrt{\frac{24n(n-1)^2}{(n-3)(n-2)(n+3)(n+5)}} \tag{5}$$

$$U_2 = g_2/\sigma_2 \tag{6}$$

在 SAS 和 SPSS 软件中均无专用的模块或程序可以直接调用以进行矩法正态性检验,但是均可通过编写程序或二次运算完成矩法正态性检验。

在 SPSS 软件中进行矩法正态性检验

在 SPSS 软件中 Descriptive Statistics(描述性统计)模块的 Frequencies(频率)过程、Descriptives(描述)过程和 Explore(探索)过程均能输出资料的偏度系数与峰度系数及其各自的标准误,然后计算得到统计量 U_1 和 U_2 ,做出正态性判定^[2]。我们在一次样本例数为 68 人的临床研究中,获得了受试者的多项检查指标,可用该方法进行正态性判定。SPSS 软件操作过程:

Analyze→Descriptive statistics→Descriptives
Dependent 框:AST,RBC,CHO
Options 框:Skewness(偏度)、Kurtosis(峰度)
OK

结果见表 1。

表 1 SPSS 软件得到的各指标的偏度系数和峰度系数

	N	偏度		峰度	
		统计量	标准误	统计量	标准误
AST	68	0.952	0.291	1.584	0.574
RBC	68	0.668	0.291	0.202	0.574
CHO	68	0.603	0.291	0.800	0.574
有效的 N(列表状态)	68				

用矩法对指标 AST 的偏度和峰度进行检验:
($U_{0.05} = 1.96$)

$$U_1 = g_1/\sigma_1 = 0.952/0.291 = 3.271 > U_{0.05}; U_2 = g_2/\sigma_2 = 1.584/0.574 = 2.760 > U_{0.05}$$

可以认为指标 AST 的偏度和峰度均不满足正态性的要求,不能认为 AST 来自正态性总体,与直方图(图 1)的结果相符,指标 AST 的分布为正偏态、尖峭峰。

分别用矩法对指标 RBC 和 CHO 的偏度和峰度进行检验:($U_{0.05} = 1.96$)

$$RBC: U_1 = g_1/\sigma_1 = 0.668/0.291 = 2.296 > U_{0.05}; U_2 = g_2/\sigma_2 = 0.202/0.574 = 0.352 < U_{0.05}$$

$$CHO: U_1 = g_1/\sigma_1 = 0.603/0.291 = 2.072 > U_{0.05}; U_2 = g_2/\sigma_2 = 0.800/0.574 = 1.394 < U_{0.05}$$

可以认为指标 RBC 和 CHO 的峰度满足正态性要求,但偏度不满足正态性的要求,为正偏态分布,也与

直方图的结果相符。

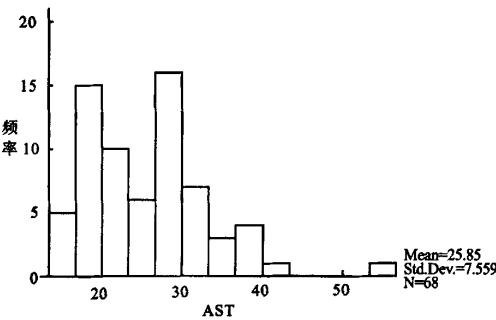


图 1 AST 的直方分布图

在 SAS 软件中进行矩法正态性检验

在 Base SAS 软件中的 means 和 univariate 两个过程均可输出资料的偏度系数和峰度系数,但是两个过程均不提供偏度系数和峰度系数的标准误选项,可以根据公式(3)~(6)编写程序输出标准误及检验统计量 U 值和概率 P 值^[6]。另外在 SAS/STAT 软件中, CALIS 过程可计算单变量和多变量的峰度^[7]。如:用 SAS 对变量 AST 进行矩法正态性检验的程序见表 2。

表 2 矩法正态性检验的 SAS 程序

行号	语句
01	data normal;
02	input ast rbc cho@@@;cards;
03	28 3.82 148
04
70	16 5.98 221
71	;
72	proc means n skewness kurtosis; var ast rbc cho;
73	output out = normal2 n = astn skewness = asts kurtosis = astk;
74	run;
75	data normal3; set normal2;
76	ast1 = (astn - 2) * (astn + 1) * (astn + 3);
	ast2 = 6 * astn * (astn - 1) / ast1;
77	astse = sqrt(ast2); uasts = asts / astse;
	Pasts = 2 * (1 - probnorm(uasts));
78	ast6 = (astn - 3) * (astn - 2) * (astn + 3) * (astn + 5);
	ast7 = 24 * astn * (astn - 1) * (astn - 1) / ast6;
79	astke = sqrt(ast7); uastk = astk / astke;
	Pastk = 2 * (1 - probnorm(uastk));
80	run;
81	proc print; var astn uasts Pasts uastk Pastk; run;

SAS 程序 01~71 行输入数据,建立数据集 'normal';72~74 行调用 means 过程对 3 个指标进行描述性统计,并将指标 AST 的样本量 astn、偏度系数 asts 和峰度系数 astk 输出到数据集 'normal2';75~80 行调用数据集 'normal2'并编程计算指标 AST 的偏度系数标准误 astse、偏度系数检验统计量 uasts、概率值 Pasts 及峰度系数标准误 astke、峰度系数检验统计量 uastk、概率值 Pastk;最后调用 print 过程输出 astn、uasts、Pasts、uastk 和 Pastk。

与直方分布图的结果相符,指标 AST 的偏度和峰度均不满足正态性的要求,不能认为 AST 来自正态性总体,指标 AST 的分布为正偏态、尖峭峰(表 3)。

同样的原理可以得到指标 RBC 和 CHO 的矩法正态性检验结果,与 SPSS 软件的结果相同。

表 3 用 SAS9.1 软件对指标 AST 进行矩法正态性检验的结果

指标	样本量	偏度系数		峰度系数	
		U 值	P 值	U 值	P 值
AST	68	3.27486	0.001057	2.75966	0.00579
RBC	68	2.29590	0.021682	0.35182	0.72498
CHO	68	2.07462	0.038022	1.39290	0.16365

矩法应用时的注意事项

当样本量很小时(比如 $n < 20$),和其他正态性检验方法一样,矩法也容易得出 $P > 0.05$ (正态性)的结果,使得犯第二类错误的概率增大,所以进行矩法正态性检验也要保证足够的样本量。

总之,矩法是进行正态性检验最恰当的方法,比其他正态性检验方法的结果更可靠,更接近直方分布图和正态概率图,我们可以通过 SPSS 软件或 SAS 软件经过编程运算来完成矩法正态性检验。

参 考 文 献

1. 郭祖超主编. 医用数理统计方法. 第 3 版. 北京:人民卫生出版社, 1988, 62-63.
2. 张文彤主编. SPSS11 统计分析教程(基础篇). 北京:北京希望电子出版社, 2002, 198-204.
3. 唐军主编. 中国医学统计百科全书·单变量推断统计分册. 北京:人民卫生出版社, 2004, 10.
4. 杨树勤主编. 中国医学百科全书·医学统计学. 上海:上海科学技术出版社, 1982, 124-129.
5. 盛骤, 谢式千, 潘承毅编. 概率论与数理统计. 北京:高等教育出版社, 2001, 250-253.
6. 高惠璇等编译. SAS 系统 Base SAS 软件使用手册. 北京:中国统计出版社, 1997, 352-359.
7. 高惠璇等编译. SAS 系统 SAS/STAT 软件使用手册. 北京:中国统计出版社, 1997, 76-79.

运用医院工作效率综合评价指数评价医院工作效率*

苏 华¹ 刘近安² 曹秀玲^{1△}

本文采用医院工作效率综合评价指数对山东省某三级甲等医院的工作效率进行统计分析,结果如下。

资料与方法

数据资料来源于山东省某三级甲等医院 1997~2005 年的统计报表,见表 1。医院工作效率综合评价指数计算公式如下:

表 1 某三级医院 1997~2005 年工作效率综合指数

年度	Q	S	R	W	P
1997	0.92	335.80	16.22	19.37	1.18
1999	0.93	339.45	15.68	18.69	1.41
2001	0.91	332.15	14.26	21.07	1.52
2003	0.90	328.50	13.55	21.44	1.55
2005	0.98	357.70	13.10	23.34	1.81

$$P = \frac{QS + W(R_0 - R)}{KQ_0}$$

其中:Q 为实际床位使用率(%), Q_0 为三级甲等医院标准

的床位使用率($Q_0 \geq 93\%$, 公式中 $Q_0 = 93\%$), S 为平均床位工作日数, R 出院者平均住院天数, R_0 三级甲等医院标准的平均住院日数($R < 23$, 公式中 $R = 23$), W 床位周转次数, K 为常数,计算年工作效率时 $K = 345$ 。

结果与分析

1997 年, P 值为 1.18, 1999 年为 1.41, 2001 年为 1.52, 2003 年为 1.55, 2005 年为 1.81。5 年的工作效率完成情况都在 100% 以上,与三级甲等医院标准相比,该医院的工作效率较高,且 2005 年比 2003 年上升了 16.77%。可见该医院工作效率在三级甲等医院中是较高的,且工作效率呈现出逐年上升的趋势。

在当今医疗市场竞争日趋激烈的现状下,为不断提高医院的工作效率,满足人民群众日益增长的医疗卫生服务的需要,该院始终坚持以病人为中心和以医疗质量为核心的医疗原则,注重人才培养,陆续添置和引进了大批先进医疗仪器设备,医院整体技术力量不断提高。尤其是 2003 年以来,该医院在狠抓质量管理的同时,加强了对人才的培养,加大了先进仪器设备的投入,促使 2005 年工作效率比 2003 年上升了 16.77%,为 1997 年以来的最高上升幅度。

* 基金项目:山东省软科学研究计划项目(2007RKA046)
1. 山东大学齐鲁医院(250012)
2. 山东大学公共卫生学院社会学与卫生事业研究所(250012)
△通讯作者:曹秀玲

作者: [徐涛](#), [韩少梅](#)
作者单位: [中国医学科学院基础医学研究所北京协和医学院基础学院统计学教研室, 100005](#)
刊名: [中国卫生统计](#) [ISTIC](#) [PKU](#)
英文刊名: [CHINESE JOURNAL OF HEALTH STATISTICS](#)
年, 卷(期): 2008, 25(3)
被引用次数: 0次

参考文献(7条)

1. [郭祖超](#) [医用数理统计方法](#) 1988
2. [张文彤](#) [SPSS11统计分析教程\(基础篇\)](#) 2002
3. [唐军](#) [中国医学统计百科全书·单变量推断统计分册](#) 2004
4. [杨树勤](#) [中国医学百科全书·医学统计学](#) 1982
5. [盛骤](#), [谢式千](#), [潘承毅](#) [概率论与数理统计](#) 2001
6. [高惠璇](#) [SAS系统Base SAS软件使用手册](#) 1997
7. [高惠璇](#) [SAS系统SAS/STAT软件使用手册](#) 1997

本文链接: http://d.g.wanfangdata.com.cn/Periodical_zgwstj200803028.aspx

授权使用: 南昌大学图书馆(wfncdxtsg), 授权号: e308617e-feb-4c51-a360-9e2b010b5eb2

下载时间: 2010年11月11日