文章编号:1007-2985(2011)03-0019-03

二维正态总体相关性的假设检验*

宋立新
(吉林师范大学博达学院,吉林 四平 136000)

摘　要:在样本相关系数具有渐近正态性的条件下,给出了一种二维正态总体分量相关性的假设检验方法,并且与大样本情形下的传统检验方法比较了功效,得到渐近相对效率为 1 的结论,且新的检验统计量形式明显简单,从而说明新的检验方法是简便可行的.

关键词:正态总体;相关性;假设检验;渐近相对效率
中图分类号:O212　　文献标志码:A

设有二维正态总体 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,其中 $\mu_1,\mu_2,\sigma_1^2>0,\sigma_2^2>0,\rho$ 均为未知参数. 欲检验原假设 $H_0:\rho=0$ VS 备择假设 $H_1:\rho\neq 0$. 由于 $(X,Y)\sim N(\cdot)$,因此 $\rho=0$ 的充要条件为 X 与 Y 独立,所以在实际中研究这类问题具有非常重要的意义.

以往,在一元线性回归分析中,可以利用皮尔逊样本矩相关系数检验,而且这种检验与回归效果的方差分析和回归系数的 t 检验是等价的[1].

笔者在此基础上首先给出大样本情形下传统的皮尔逊样本矩相关系数检验法,然后在样本相关系数具有渐近正态性的条件下,构造了一种二维正态总体分量相关性的假设检验方法,并且与大样本情形下的传统检验方法比较了功效.

1　相关引理

设总体 (X,Y) 服从二元正态分布 $N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$,其密度函数为 $f(x,y)=\dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\{-\dfrac{1}{2(1-\rho^2)}[\dfrac{(x-\mu_1)^2}{\sigma_1^2}-\dfrac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\dfrac{(y-\mu_2)^2}{\sigma_2^2}]\}$. 从总体抽取容量为 n 的样本 $(X_1,Y_1),(X_2,Y_2),\cdots,(X_n,Y_n)$,定义样本相关系数为

$$r_n=\frac{\sum_{i=1}^n(X_i-\bar X)(Y_i-\bar Y)}{\sqrt{\sum_{i=1}^n(X_i-\bar X)^2\sum_{i=1}^n(Y_i-\bar Y)^2}},\bar X=\frac{1}{n}\sum_{i=1}^n X_i,\bar Y=\frac{1}{n}\sum_{i=1}^n Y_i.$$

引理 1　根据文献[2],样本相关系数满足 $\dfrac{r_n}{\sqrt{1-r_n^2}}=\dfrac{\frac{\rho}{\sqrt{1-\rho^2}}t_{11}+t_{21}}{t_{22}}$,其中 $t_{11}^2\sim\chi^2(n-1),t_{22}^2\sim$

* 收稿日期:2011-03-18
基金项目:国家自然科学基金资助项目(10971084);吉林省教育厅“十一五”科学技术重点研究项目(吉教科合字[2010]-141)
作者简介:宋立新(1954-),男,吉林乾安人,吉林师范大学博达学院教授,主要从事 Bayes 统计研究.

$\chi^2(n-2)$,$t_{21}\sim N(0,1)$,且 t_{11},t_{22},t_{21} 相互独立.特别地,当 $\rho=0$ 时,$\dfrac{r_n}{\sqrt{1-r_n^2}}=\dfrac{t_{21}}{t_{22}}$,所以$\dfrac{r_n}{\sqrt{1-r_n^2}}\sqrt{n-2}\sim t(n-2)$.

引理 2 根据文献[3],样本相关系数$\dfrac{\sqrt{n}(r_n-\rho)}{1-\rho^2}\xrightarrow{L}N(0,1)$.

引理 3 $\dfrac{\chi(n-1)}{\chi(n-2)}\xrightarrow{P}1$.

证明 先证明$\dfrac{\chi^2(n-1)}{\chi^2(n-2)}\xrightarrow{P}1$.根据柯赫伦定理,$\chi^2(n-1)$ 可以分解成$\sum\limits_{i=1}^{n-1}Z_i^2$,$\chi^2(n-2)$ 可以分解成$\sum\limits_{i=1}^{n-2}Z_i^2$,其中 $Z_i\sim N(0,1)$,且 $Z_1,Z_2,\cdots,Z_{n-1}$ 相互独立.因为 $E\chi^2(n-1)=n-1$,$E\chi^2(n-2)=n-2$.根据辛钦大数定律,$\dfrac{\chi^2(n-1)}{n-1}\xrightarrow{P}1$,$\dfrac{\chi^2(n-1)}{n-1}\xrightarrow{P}1$ 且$\sqrt{\dfrac{n-2}{n-1}}\to 1(n\to\infty)$.根据 Slutsky 定理,有$\dfrac{\chi^2(n-1)}{\chi^2(n-2)}=\dfrac{\chi^2(n-1)/(n-1)}{\chi^2(n-2)/(n-2)}\cdot\dfrac{n-1}{n-2}\xrightarrow{P}1$,从而$\dfrac{\chi(n-1)}{\chi(n-2)}\xrightarrow{P}1$.

2 大样本情形下的传统检验

传统的皮尔逊样本矩相关系数检验法如下:欲检验 $H_0:\rho=0$ VS $H_1:\rho\neq 0$.先求出样本相关系数,对于给定的显著性水平 α,查相关系数临界值表有 $r_{1-\alpha}$[1].若 $|r_n|\geqslant r_{1-\alpha}$,否定 H_0.若 $|r_n|<r_{1-\alpha}$,不能否定 H_0.在 H_0 成立且 n 很大时,根据引理 1 和 t 分布的极限分布为标准正态分布的结论,有$\dfrac{r_n}{\sqrt{1-r_n^2}}\sqrt{n-2}\sim N(0,1)$.对于给定显著性水平 α,查 $N(0,1)$ 表,取 $U_{1-\alpha/2}$.根据$\dfrac{r_n}{\sqrt{1-r_n^2}}\sqrt{n-2}$ 关于 r_n 在 $-1\leqslant r_n\leqslant 1$ 上单调递增的性质,若 $\left|\dfrac{r_n}{\sqrt{1-r_n^2}}\sqrt{n-2}\right|\geqslant U_{1-\alpha/2}$,否定 H_0;若 $\left|\dfrac{r_n}{\sqrt{1-r_n^2}}\sqrt{n-2}\right|<U_{1-\alpha/2}$,不能否定 H_0.

3 新的检验方法

欲检验 $H_0:\rho=0$ VS $H_1:\rho\neq 0$.当 H_0 成立时,根据引理 2,有$\sqrt{n}r_n\xrightarrow{L}N(0,1)$.用$\sqrt{n}r$ 作为检验统计量,那么对于给定的显著性水平 α,查 $N(0,1)$ 表,有 $U_{1-\alpha/2}$,使得 $P(|\sqrt{n}r_n|\geqslant U_{1-\alpha/2})=\alpha$.若 $|\sqrt{n}r_n|\geqslant U_{1-\alpha/2}$,否定 H_0,认为 X 与 Y 相关;否则,当 $|\sqrt{n}r_n|<U_{1-\alpha/2}|$ 时,否定 H_1,认为 X 于 Y 不相关.类似的方法可以得到另外 2 种单侧检验.

4 2 种检验的渐近相对效率

接下来考虑 2 种检验的渐近相对效率[4].为方便起见,先考虑备择假设是单侧假设情形.令 $H_0:\rho=0$ VS $H_1:\rho>0$;另一种单侧检验和双侧检验可用类似方法求得.

(1) 在大样本情形下,传统检验要达到检验功效 β 时必需的样本容量 n_s.

$$S_n=\frac{r_n}{\sqrt{1-r_n^2}}\sqrt{n-2}=\frac{\dfrac{\rho}{\sqrt{1-\rho^2}}t_{11}+t_{21}}{t_{22}}\sqrt{n-2}.$$

在 H_0 成立且 n 很大时,水平为 α 的检验拒绝域近似为$\{S_n\geqslant U_{1-\alpha/2}\}$.在 H_1 成立时计算功效的参数 ρ 要与样本容量 n 有关,且样本容量 n 越大,ρ 越接近 0.为此,取一个参数序列 $\rho_1\geqslant\rho_2\geqslant\cdots\geqslant\rho_n\geqslant\cdots>0$,其中 $\rho_n=\dfrac{\delta}{\sqrt{n}}$,$\delta$ 是大于 0 的正数,显然在 $n\to\infty$ 时,$\rho_n\to 0$.所谓在 ρ_n 处计算功效,就是在备择假设成立,相关系数为 ρ_n 时计算功效,即计算概率 $P(S_n\geqslant U_{1-\alpha})$ 的值.而对于 S_n,由引理 1,$\dfrac{t_{21}}{t_{22}}\sqrt{n-2}\sim t(n-2)$,则$\dfrac{t_{21}}{t_{22}}$

$\sqrt{n-2} \xrightarrow{L} N(0,1)$，$\dfrac{t_{11}}{t_{22}} \xrightarrow{P} 1$，$\sqrt{1-\rho^2} \to 1$，$\dfrac{\sqrt{n-2}}{\sqrt{n}} \to 1(n \to \infty)$. 根据 Slutsky 定理，$S_n \xrightarrow{L} N(\delta,1)$. 所以功效 $P(S_n \geqslant U_{1-\alpha}) = \beta$，有 $U_{1-\alpha} - \delta = U_{1-\beta}$. 一般来说，水平 α 比较小时，功效 β 比较大，故 $U_{1-\alpha} > 0$ 而 $U_{1-\beta} < 0$，$\delta = U_{1-\alpha} - U_{1-\beta}$. 令样本容量为 n_s，有 $\rho = \dfrac{\delta}{\sqrt{n_s}}$，则 $\sqrt{n_s}\rho = \delta = U_{1-\alpha} - U_{1-\beta}$，从而 $n_s = \dfrac{(U_{1-\alpha} - U_{1-\beta})^2}{\rho^2}$.

(2) 新的检验要达到检验功效 β 时必需的样本容量 n_t.

检验统计量 $t_n = \sqrt{n}\dfrac{r_n - \rho}{1-\rho^2}$，由引理 2，在原假设 H_0 成立时，$\rho = 0$，$t_n = \sqrt{n}r_n \xrightarrow{L} N(0,1)$. 水平为 α 的检验的拒绝域近似为 $\{t_n \geqslant U_{1-\alpha}\}$. 在 H_1 成立时 ρ_n 的取法同(1). 由引理 2 知 $\sqrt{n}\dfrac{(r_n - \rho_n)}{1-\rho_n^2} \xrightarrow{L} N(0,1)$. 由于 $\sqrt{n}\rho_n = \delta$，在 $n \to \infty$ 时，$1-\rho_n^2 \to 1$，因此 $\sqrt{n}\dfrac{(r_n - \rho_n)}{1-\rho_n^2}(1-\rho_n^2) + \sqrt{n}\rho_n \xrightarrow{L} N(\delta,1)$，$n \to \infty$. 那么功效 $\beta = P(\sqrt{n}r_n \geqslant U_{1-\alpha}) = P(\sqrt{n}\dfrac{(r_n - \rho_n)}{1-\rho_n^2}(1-\rho_n^2) + \sqrt{n}\rho_n \geqslant U_{1-\alpha})$. 即 $P(N(\delta,1)_n \geqslant U_{1-\alpha})$，从而 $U_{1-\alpha} - \delta = U_{1-\beta}$，$\delta = U_{1-\alpha} - U_{1-\beta}$. 令样本容量为 n_t，令 $\rho = \dfrac{\delta}{\sqrt{n_t}}$，则 $\sqrt{n_t}\rho = \delta = U_{1-\alpha} - U_{1-\beta}$，从而 $n_t = \dfrac{(U_{1-\alpha} - U_{1-\beta})^2}{\rho^2}$.

比较 n_s 和 n_t 的大小，当 $n \to \infty$，$\dfrac{n_s}{n_t} \to 1$，即 2 种方法的渐近相对效率为 1；加之，在 H_0 成立时 $t_n = \sqrt{n}r_n$ 明显比 S_n 简单：这充分说明新的检验法是简便可行的.

注 1　因为 2 个正态总体在相关时并不一定能产生二维正态总体，所以文中只限于考虑二维正态总体 2 个分量相关性. 但新的检验法是在原假设 $H_0: \rho = 0$ 成立时得到的，而在总体联合正态条件下 $\rho = 0$ 当且仅当两分量 X 与 Y 独立，因此该方法也可用于检验 2 个正态总体是否独立，只不过求渐近相对效率时未必用文中的方法.

参考文献:

[1] 茆诗松，程依明，濮晓龙. 概率论与数理统计 [M]. 北京：高等教育出版社，2004：400.

[2] 张尧庭，方开泰. 多元统计分析引论教程 [M]. 北京：科学出版社，1997：488-490.

[3] 赵志文，刘银萍，宋立新. 二元正态总体样本相关系数的渐近正态性 [J]. 佳木斯大学学报，2009(3)：182-184.

[4] 王静龙，梁小筠. 非参数统计分析 [M]. 北京：高等教育出版社，2006：152-162.

Hypothesis Testing of the Dependency of Two-Dimensional Normal Population

SONG Li-xin

(College of Boda，Jilin Normal University，Siping 136000，Jilin China)

Abstract：A hypothesis testing of the dependency of two-dimensional normal population is given，under the condition of the sample correlation coefficient with asymptotic normality. Compared with the traditional testing method the asymptotic relative efficiency is 1. The new test statistic is simple，and the new testing method is proved to be easy and feasible.

Key words：normal population；dependency；hypothesis teating；asymptotic relative efficiency

(责任编辑　向阳洁)