·计算机应用·

用 t 检验比较多个均数犯Ⅰ类错误概率的 SAS电脑实验

厦门大学医学院(361005) 韩耀风 方 亚△

在“医学统计学”课程教学中,对于多个均数的比较能否用 t 检验,往往是通过讨论说明直接用 t 检验可增大犯Ⅰ类错误的概率,从而引出方差分析的方法。针对这一部分内容的教学,如果辅以相应的电脑实验则可帮助学生理解“多个均数的比较不宜用 t 检验代替”的缘由。然而在现有的教材里,主要存在两个问题:(1)犯Ⅰ类错误的判定不明确;(2)得出的犯Ⅰ类错误的概率不符合概率的统计定义[1]。另外,在进行 SAS电脑实验时,总体均数和标准差以及组数是否会影响模拟计算的结果?在模拟计算时如何设定这些参数?针对这两个问题,如何正确应用电脑实验以及如何选择参数进行模拟计算,以解决“多个均数比较不宜用 t 检验代替”的教学实际问题,本文将从Ⅰ类错误的定义出发,对多个均数比较的Ⅰ类错误进行判定,从而给出理论上用 t 检验进行多个均数比较所犯Ⅰ类错误的概率,并进行 SAS模拟计算。通过比较模拟值与设定的检验水准,探讨上述 SAS电脑实验的最佳模拟条件,同时进行 Bonferroni校正后犯Ⅰ类错误概率的 SAS电脑实验,以加深学生对“多个均数的比较不宜用 t 检验代替”的印象,从而提高学生正确选择统计方法的思维和应用能力。

方 法

1. Ⅰ类错误的定义

如果实际情况与零假设 H_0 一致,由于抽样的原因使得统计量的观察值落到拒绝域,拒绝原本正确的 H_0,导致推断结论错误,该错误称为Ⅰ类错误[2,3]。

多个均数比较的 H_0 为: k 个总体均数相等; H_1: k 个总体均数不全相等。若用 t 检验进行两两比较需要比较 $k(k-1)/2$ 次,若 $k(k-1)/2$ 次 t 检验结果均为 $P>\alpha$,则不能拒绝 H_0,若至少有一次 t 检验的结果为 $P\leqslant\alpha$,则可以拒绝 H_0。由Ⅰ类错误的定义可知,若实际情况为 k 个总体均数相等, $k(k-1)/2$ 次 t 检验中至少有一次 t 检验的结果为 $P\leqslant\alpha$,则犯Ⅰ类错误。

两个均数比较 t 检验的 H_0 为:两个总体均数相等; H_1:两个总体均数不相等。若实际情况为两个总体均数相等, t 检验的结果为 $P\leqslant\alpha$,则犯Ⅰ类错误。

多个均数比较的Ⅰ类错误与 t 检验的Ⅰ类错误的关系为:若 $k(k-1)/2$ 次 t 检验均不犯Ⅰ类错误,则多个均数比较不犯Ⅰ类错误;若 $k(k-1)/2$ 次 t 检验中至少有一次犯Ⅰ类错误,则多个均数比较犯Ⅰ类错误。

2 用 t 检验进行多个均数比较所犯Ⅰ类错误概率的理论值

已知一次 t 检验犯Ⅰ类错误的概率为 α,则一次 t 检验不犯Ⅰ类错误的概率为 $1-\alpha$,根据独立事件概率的乘法原则[4], $k(k-1)/2$ 次 t 检验均不犯Ⅰ类错误的概率为 $(1-\alpha)^{k(k-1)/2}$,则 $k(k-1)/2$ 次 t 检验中至少有 1次犯Ⅰ类错误的概率为 $1-(1-\alpha)^{k(k-1)/2}$,即用 t 检验进行多个均数比较时犯Ⅰ类错误的概率为 $1-(1-\alpha)^{k(k-1)/2}$。假定 $\alpha=0.05$, $k=5$,则 $1-(1-\alpha)^{k(k-1)/2}=0.4013>>0.05$。

3. Bonferroni校正

采用 Bonferroni校正的要点是调整检验水准[2]。若每次 t 检验的检验水准调整为 $\alpha'=\frac{\alpha}{k(k-1)/2}$,则犯Ⅰ类错误的累积概率不超过 α[2],即用 t 检验进行多个均数比较时犯Ⅰ类错误的概率不超过 α。按照前面的推论,容易得出若对 t 检验进行 Bonferroni校正后犯Ⅰ类错误概率的理论值为 $1-(1-\alpha')^{k(k-1)/2}$。假定 $\alpha=0.05$, $k=5$,则 $1-(1-\alpha')^{k(k-1)/2}=0.0489<0.05$,说明应用 Bonferroni校正可使多个均数比较的假阳性率不超过设定的检验水准。

4. SAS宏程序的编制

用 t 检验进行多个均数的比较所犯Ⅰ类错误概率的 SAS电脑实验如下:设组数为 k,各组的样本量均为 n, k 个样本均来自正态总体 $N(\mu,\sigma^2)$,抽样后共进行 $k(k-1)/2$ 次 t 检验,检验水准定为 0.05,若至少有一次的结果为 $P\leqslant0.05$,则认为犯Ⅰ类错误,否则没有犯Ⅰ类错误。根据概率的统计定义[1]:设在相同条件下,进行大量独立重复试验,若事件 A 的频率稳定地在某一确定值 p 附近摆动,称此数值 p 为事件 A 发生的概率,记作 $P(A)=p$。本实验共重复试验 2 000次,统计犯Ⅰ类错误的次数为 q,则犯Ⅰ类错误的概率为 $\frac{q}{2000}\times100\%$。

Bonferroni校正后犯Ⅰ类错误概率的 SAS电脑实

△通讯作者:方亚, E-mail: fangya@xmu.edu.cn

验如下:实验数据同上。若 $k(k-1)/2$次 t检验中至少有一次结果为 $P\leqslant\frac{0.05}{k(k-1)/2}$,认为犯Ⅰ类错误,否则没有犯Ⅰ类错误。共重复试验 2 000次,统计犯Ⅰ类错误的次数为 q',则犯Ⅰ类错误的概率为 $\frac{q'}{2000}\times 100\%$。

为方便应用,作者编制了 SAS宏程序,应用时只修改有关的参数即可。SAS宏程序如下:

```
options nosource nonotes;
%macro type1(repli=, alpha=, g=, n0=, mean=, sd=, seed0=);
/* repli=定义试验的重复次数, alpha=所设检验水准, n0=定义每组的样本量, g=定义被比较的组数, mean=定义总体均数, sd=定义总体标准差, seed0=定义的种子数 */
data b; run;
%do s=1 %to &repli;
data a;
k=&g; n=&n0; seed=&seed0* &s; mu=&mean; sigma=&sd;
do grp=1 to k;
do no=1 to n;
x=rannor(seed)* sigma+mu; output;
end;
end;
proc means noprint;
var x;
by grp;
output out=a_0 mean=mean var=var;
data a_5;
set a_0;
keep mean;
proc transpose out=a_2 prefix=m;
data a_3;
set a_0; keep var;
proc transpose out=a_4 prefix=v;
data aall;
merge a_2 a_4;
array m(&g); array v(&g);
k=&g; n=&n0;
do i=1 to k-1;
do j=i+1 to k;
t=(m(i)-m(j)) /sqrt((v(i)+v(j))/n);
p=(1-probt(abs(t),2* n-2))* 2;
if p<=&alpha then count+1;
end;
end;
c=(count>=1);
data b; set b aall; run;
%end;
proc means data=b sum mean std n;
var c;
run;
%mend;
```

例如,试验重复次数为 2 000,检验水准为 0.05,组数为 3,每组样本量为 10,各组总体均数为 0,总体标准差为 1,种子数为 1 000时,首先运行上面的宏程序,然后可运行如下 SAS语句:

```
%type1(repli=2000, alpha=0.05, g=3, n0=10, mean=0, sd=1, seed0=1000);
```

SAS输出结果中,变量 c的均值即为用 t检验进行多个均数比较所犯Ⅰ类错误的概率。

结　果

在不同组数、不同样本量及不同总体均数和标准差情况下,用 t检验进行多个均数的比较所犯Ⅰ类错误概率的理论值及模拟值见表 1。可见,在组数和样本量不变的情况下,犯Ⅰ类错误概率的模拟值不随总体均数和标准差的改变而改变;在组数、总体均数和标准差不变的情况下,该模拟值受样本量的影响甚微;在样本量、总体均数和标准差不变的情况下,组数越大,该模拟值越大,且均大于设定的检验水准 0.05。

采用 Bonferroni校正后,犯Ⅰ类错误概率的模拟值比校正前低,且小于原检验水准 0.05。

表 1　用多次 t检验进行多个均数比较的Ⅰ类错误概率

	$\alpha=0.05$		$\alpha=0.05$/比较次数(Bonferroni校正)	
	理论值*	模拟值	理论值*	模拟值
$k=3, n=10, \bar{x}=0, S=1$	0.1426	0.1225	0.0492	0.0380
$k=3, n=20, \bar{x}=0, S=1$		0.1220		0.0395
$k=3, n=30, \bar{x}=0, S=1$		0.1340		0.0460
$k=3, n=10, \bar{x}=10, S=5$		0.1225		0.0380
$k=3, n=20, \bar{x}=10, S=5$		0.1220		0.0395
$k=3, n=30, \bar{x}=10, S=5$		0.1340		0.0460
$k=5, n=10, \bar{x}=0, S=1$	0.4013	0.2955	0.0489	0.0365
$k=5, n=20, \bar{x}=0, S=1$		0.2935		0.0455
$k=5, n=30, \bar{x}=0, S=1$		0.2970		0.0410
$k=5, n=10, \bar{x}=10, S=5$		0.2955		0.0365
$k=5, n=20, \bar{x}=10, S=5$		0.2935		0.0455
$k=5, n=30, \bar{x}=10, S=5$		0.2970		0.0410
$k=10, n=10, \bar{x}=0, S=1$	0.9006	0.6265	0.0488	0.0430
$k=10, n=20, \bar{x}=0, S=1$		0.6310		0.0335
$k=10, n=30, \bar{x}=0, S=1$		0.6400		0.0350
$k=10, n=10, \bar{x}=10, S=5$		0.6265		0.0430
$k=10, n=20, \bar{x}=10, S=50$		0.6310		0.0335
$k=10, n=30, \bar{x}=10, S=5$		0.6400		0.0350

*:为直接用 t检验进行多个均数的比较所犯Ⅰ类错误概率的理论值由 $1-(1-\alpha)^{k(k-1)/2}$推论得出。

讨　　论

医学统计学是一门处理医学研究数据中变异性的科学与艺术, 其目的是获得可靠的结果。在进行“医学统计学”课程教学中, 培养学生的统计学思维能力和正确选择统计方法是非常重要的环节, 例如能否用所学 t检验进行多个均数的比较的问题。笔者通过多年的教学实践发现医学院的学生大部分缺乏概率论基础, 对于能否用所学 t检验进行多个均数的比较的问题很难从理论上理解这一部分内容, 教学时, 往往通过讨论直接用 t检验可增大犯Ⅰ类错误的概率, 从而引出方差分析的方法。如果辅以相应的电脑实验可帮助学生理解“多个均数的比较不宜直接用 t检验代替”的缘由, 也可以帮助学生在实际应用中正确选择统计方法。

用 t检验进行多个均数的比较所犯Ⅰ类错误概率的模拟计算能否正确显示主要取决于两个因素: Ⅰ类错误的正确判定和是否通过大量的独立重复实验估计概率。文中提供的电脑实验中犯Ⅰ类错误的判定是建立在Ⅰ类错误的定义基础上, 是正确的、合理的。根据概率的统计定义, 本文提供的电脑实验在相同条件下重复 2 000次, 估计犯Ⅰ类错误的概率为 2 000次试验中发生Ⅰ类错误的频率, 符合概率的统计定义。

从模拟结果可以看出, 直接用 t检验进行多个均数比较犯Ⅰ类错误的概率与总体的均数和标准差没有关系, 而与组数有关。在 SAS电脑实验中, 正态分布随机数的模拟按如下方法进行: 首先模拟出标准正态分布的随机数 z, 然后对 z进行变换 (z × sd+ mean)生成任意正态分布的随机数。SAS中正态分布随机数的生成方式导致模拟计算的结果与总体的均数和标准差没有关系; 样本量不变, 组数越多, 犯Ⅰ类错误的概率越高; 组数不变, 犯Ⅰ类错误的概率受样本量的影响甚微。那么在今后的教学中, 应该如何选择模拟条件? 本文电脑实验的目的是验证用 t检验进行多个均数的比较可增大犯Ⅰ类错误概率, 那就要求模拟实验得出的用 t检验进行多个均数比较犯Ⅰ类错误的概率要大于所设定的检验水准, 而用 Bonferroni校正后的 t检验犯Ⅰ类错误的概率不超过设定的检验水准。模拟实验结果显示, 用 t检验获得的模拟值均大于设定的检验水准, 而采用 Bonferroni校正后的 t检验的模拟值均小于设定的检验水准。由此可见, 在“用 t检验进行多个均数比较犯Ⅰ类错误概率”的电脑实验中模拟参数可以任意设定。

本文比较了直接用 t检验和 Bonferroni校正后进行多个均数比较犯Ⅰ类错误的概率, 得出在同等条件下即相同组数、相同样本量、相同的总体均数和标准差, Bonferroni校正后可降低犯Ⅰ类错误的概率, 从而帮助学生更好地理解为什么不宜直接用 t检验进行多个均数的比较而可以用 Bonferroni校正的方法进行两两比较。

小　　结

(1)本文提供了一个可以正确模拟显示“用 t检验进行多个均数比较将增大犯Ⅰ类错误的概率”的电脑实验。

(2)在“用 t检验进行多个均数比较犯Ⅰ类错误概率”的电脑实验中的模拟参数可以任意设定。

(3)培养学生的统计学思维能力和正确选择统计方法是“医学统计学”课程教学中的重要环节。

参 考 文 献

1. 祝国强, 主编. 医药数理统计方法. 北京: 高等教育出版社, 2004.
2. 方积乾主编. 卫生统计学 (第 6版). 北京: 人民卫生出版社, 2008
3. 茆诗松, 王静龙, 濮晓龙著. 高等数理统计 (第 1版). 北京: 高等教育出版社, 2008.
4. 杨洪礼, 鲍承友, 张序萍主编. 概率论与数理统计. 北京: 北京邮电大学出版社, 2007.

(上接第 540页)

疗服务水平有重要的意义。熵权模糊积分法在评价过程中通过合理确定指标所占权重, 使评价结果更符合现实情况, 该法可以作为医院评价年度或阶段医疗管理工作的一种重要手段。

参 考 文 献

1. 刘涛, 邓平基, 孟晓谕. 基于熵权法在医疗质量综合评价. 中国卫生统计, 2009, 26(3): 274-275.
2. 刘涛, 孟晓谕, 邓平基. 模糊灰色评价方法在医疗质量评价中的应用. 中国卫生统计, 2009, 26(2): 188-190
3. 冯秀娟. 应用直接综合法对医院医疗质量进行综合评价. 中国卫生统计, 2006, 23(4): 379-380
4. 高丽娟. TOPSIS法对医院医疗质量管理的综合评价. 中国卫生统计, 2007, 24(3): 300-301.