· 专题研究 ·

样本量估计及其在 nQuery 和 SAS 软件上的实现
——均数比较(三)

段重阳[1] 吕 朵[1] 陈平雁[2△]

1.2 两样本均数的比较

1.2.1 差异性检验

1.2.1.1 两样本 t 检验(方差齐性)

方法:O'Brien 和 Muller(1983)[2-3]等给出两样本 t 检验的样本量估计是建立在自由度为 $2(n-1)$,非中心参数为 $\sqrt{n}\left(\frac{\mu_1-\mu_2}{\sigma\cdot\sqrt{2}}\right)$ 的非中心 t 分布上。其检验效能的计算公式为:

$$1-\beta=1-\text{Prob}t\left(t_{1-\alpha/s,2(n-1)},2(n-1),\sqrt{n}\left(\frac{\mu_1-\mu_2}{\sigma\cdot\sqrt{2}}\right)\right) \tag{1-8}$$

式中,μ_1,μ_2 分别为两样本均数,σ 为样本标准差。

在计算样本量时,一般先设定样本量初始值,然后迭代样本量直到所得的检验效能满足条件为止。此时的样本量,即研究所需的样本量。

【例 1-7】欲研究一种新药对老年妇女红细胞比容(HCT)的影响,从已有关于 HCT 的研究报道中获得如下信息:

① 预试验结果显示,6 名髋骨骨折后的老年妇女的平均 HCT 值为 32.3%,32 名正常老年妇女的平均 HCT 值为 33.5%。

② 在这种新药对其他患病人群的临床试验中,安慰剂组的病人 HCT 没有改变,而试验组的病人 HCT 升高 2.5% ~5%。

综合上述信息,我们可以预计安慰剂组 HCT 没有变化,而试验组会有 2.5% ~5% 的升高。据此,我们保守估计新药可将老年妇女的 HCT 升高 2% ~2.4%。以往报道 HCT 的标准差在 1.5% ~2.5% 之间,因此我们用 2% 估计 HCT 的标准差。当检验效能为 90% 时,试估计样本量。

nQuery Advisor 7.0 实现:设定检验水准 $\alpha=0.05$;双侧检验,即 $s=2$;检验效能取 $1\text{-}\beta=90\%$。依据上述基础数据可知,$\mu_1=2$、2.2、2.4,$\mu_2=0$,$\sigma=2$,在 nQuery Advisor 7.0 主菜单选择:

Goal:Make Conclusion Using:⊙ Means

Number of Groups:⊙ Two

Analysis Method:⊙ Test

方法框中选择:Student's t test(equal variance)。

在弹出的样本量计算窗口将各参数键入,如图 1-17 所示,结果 n 分别为 23、19、16。

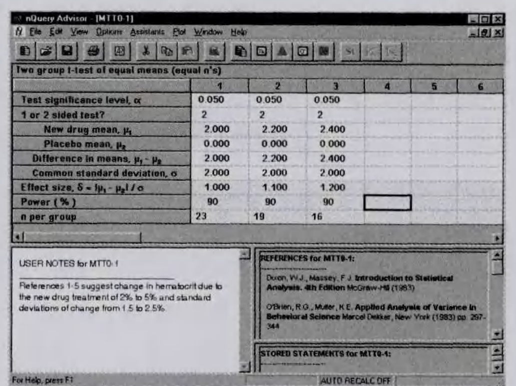

图 1-17 nQuery Advisor 7.0 关于例 1-7 样本量估计的参数设置与计算结果

从计算结果我们看到每组需要 16 到 23 例(总共需要 32 到 46 例)。

SAS9.2 程序:

```
PROC IML;
start MTT0(a,s,mean1,mean2,sd,power);
error=0;
if(a>1|a<0) then do;error=1;print "error" "Test significance level must be in 0-1";end;
if(s^=1 & s^=2) then do;error=1;print "error" "s=1 or 2";end;
if(sd<0) then do;error=1;print "error" "Standard deviation must be >=0";end;
if(power>100 | power<1) then do;error=1;print "error" "Power(%) must be in 1-100";end;
if(error=1) then stop;
if(error=0)then do;
es=abs(mean1-mean2)/sd;
n=2;
do until(pw>=power/100);
ncp=sqrt(n)#es/sqrt(2);
```

1. 南方医科大学 公共卫生与热带医学学院 生物统计学系 2007 级本科生
2. 南方医科大学 公共卫生与热带医学学院 生物统计学系
△通讯作者:陈平雁

Chinese Journal of Health Statistics, Apr 2012, Vol. 29, No. 2

```
df = 2#(n-1);
t = tinv(1-a/s, df);
pw = 1-probt(t, df, ncp);
n = n + 0.01;
end;
n = ceil(n-0.01);
print a[label = "Test Significance level"]
s[label = "1 or 2 sided test"]
mean1[label = "New drug mean"]
mean2[label = "Placebo mean"]
sd[label = "Common standard deviation"]
power[label = "Power(%)"]
n[label = "n"];
end;
finish MTT0;
run MTT0(0.05,2,2,0,2,90);
run MTT0(0.05,2,2.2,0,2,90);
run MTT0(0.05,2,2.4,0,2,90);
quit;
```

SAS 运行结果：

Test Significance level	1 or 2 sided test	New drug mean	Placebo mean	Common standard deviation	Power (%)	n
0.05	2	2	0	2	90	23
0.05	2	2.2	0	2	90	19
0.05	2	2.4	0	2	90	16

图 1－18　SAS 9.2 关于例 1－7 样本量估计的参数设置与计算结果

1.2.1.2　Satterthwaite t 检验（方差不齐）

方法：Moser 等（1989）[7] 给出 Satterthwaite t 检验样本量的估计是建立在调整自由度 v，非中心参数为 $\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$ 的非中心 t 分布上，调整自由度按下式计算：

$$v=\frac{\left(\frac{1}{n_1}+\frac{r}{n_2}\right)^2}{\frac{1}{n_1^2(n_1-1)}+\frac{r^2}{n_2^2(n_2-1)}} \qquad (1-9)$$

其检验效能的计算公式为：

$$1-\beta=1-\mathrm{Prob}t\left(t_{1-\alpha/s,v},v,\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}\right) \qquad (1-10)$$

式中，μ_1，μ_2 分别为两样本均数，σ_1，σ_2 分别为两总体标准差；$r=\sigma_2^2/\sigma_1^2$。

在计算样本量时，一般先设定样本量初始值，然后迭代样本量直到所得的检验效能满足条件为止。此时的样本量，即研究所需的样本量。

【例 1－8】假设一个平行随机对照的临床试验比较两种药物对激素水平的影响，预试验结果显示：对照组的均数和标准差分别为 300 和 150，试验组的分别为 1200 和 600。设定检验效能为 90%，采用平衡设计，试估计样本量。

nQuery Advisor 7.0 实现：设定检验水准 $\alpha=0.05$；双侧检验，即 $s=2$；检验效能取 $1-\beta=90\%$。依据上述基础数据可知，$\mu_1=300$，$\mu_2=1200$，$\sigma_1=150$，$\sigma_2=600$；

在 nQuery Advisor 7.0 主菜单选择：

Goal：Make Conclusion Using：⊙ Means

Number of Groups：⊙ Test

Analysis Method：⊙ Test

方法框中选择：Satterthwaite t-test（unequal variances）。

在弹出的样本量计算窗口将各参数键入，如图 1－19 所示，结果 $n=7$。

Two group Satterthwaite t-test of equal means (unequal variances) (equal n's)	1	2	3	4	5
Test significance level, α	0.050				
1 or 2 sided test?	2				
Group 1 mean, μ_1	300.000				
Group 2 mean, μ_2	1200.000				
Difference in means, $\mu_1-\mu_2$	-900.000				
Group 1 standard deviation, σ_1	150.000				
Group 2 standard deviation, σ_2	600.000				
Power (%)	90				
n per group	7				

图 1－19　nQuery Advisor 7.0 关于例 1－8 样本量估计的参数设置与计算结果

SAS9.2 软件实现：

```
PROC IML;
start MTT0uv(a, s, mean1, mean2, sd1, sd2, ratio, power);
error = 0;
if(a > 1 | a < 0) then do; error = 1; print "error" "Test significance level must be in 0-1"; end;
if(s^ = 1 & s^ = 2) then do; error = 1; print "error" "s = 1 or 2"; end;
if(sd1 < 0) then do; error = 1; print "error" "Standard deviation of group 1 must be > = 0"; end;
if(sd2 < 0) then do; error = 1; print "error" "Standard deviation of group 2 must be > = 0"; end;
if(ratio < 0) then do; error = 1; print "error" "n2/n1 must be > = 0 "; end;
if(power > 100 | power < 1) then do; error = 1; print "error" "Power(%) must be in 1-100"; end;
```

```
if(error = 1) then stop;
if(error = 0)then do;
sita = sd2##2/sd1##2;
n1 = 2;
do until(pw > = power/100);
n2 = ratio#n1;
ncp = abs(mean2-mean1)/sqrt(sd1##2/n1 + sd2##
2/n2);
v = (1/n1 + sita/n2)##2/(1/(n1##2#(n1-1)) +
sita##2/(n2##2#(n2-1)));
t = tinv(1-a/s,v);
pw = 1-probt(t,v,ncp);
n1 = n1 + 0.01;
end;
n1 = ceil(n1-0.01);
n2 = ceil(n1#ratio);
print a[label = "Test Significance level"]
s[label = "1 or 2 sided test"]
mean1[label = "Group 1 mean"]
mean2[label = "Group 2 mean"]
sd1[label = "Group 1 standard deviation"]
sd2[label = "Group 2 standard deviation"]
power[label = "Power(%)"]
n1[label = "n1"]
n2[label = "n2"];
end;
finish MTT0uv;
run MTT0uv(0.05,2,300,1200,150,600,1,90);
quit;
```

SAS 运行结果：

Test Significance level	1 or 2 sided test	Group 1 mean	Group 2 mean	Group 1 standard deviation	Group 2 standard deviation	Power (%)	n1	n2
0.05	2	300	1200	150	600	90	7	7

图 1-20 SAS 9.2 关于例 1-8 样本量估计的参数设置与计算结果

1.2.1.3 基于对数正态分布的两样本 t 检验

方法：根据 Diletti 等[2-3,8]的方法，先将均数和标准差进行对数转换，然后按照两样本 t 检验的样本量估计方法计算。因此，检验效能及样本量估计是建立在自由度为 $2(n-1)$，非中心参数为 $\sqrt{n}\left(\frac{\ln(FC)}{\sqrt{\ln(1+CV^2)}\cdot\sqrt{2}}\right)$的非中心 t 分布。其检验效能的计算公式为：

$$1-\beta = 1-\text{Prob}t\left(t_{1-\alpha/s,2(n-1)},\ 2(n-1),\ \sqrt{n}\left(\frac{\ln(FC)}{\sqrt{\ln(1+CV^2)}\cdot\sqrt{2}}\right)\right) \qquad (1-11)$$

式中，$\sigma=\sqrt{\ln(1+CV^2)}$，$FC$ 为较大均数和较小均数的比值；CV 为变异系数。

在计算样本量时，一般先设定样本量初始值，然后迭代样本量直到所得的检验效能满足条件为止。此时的样本量，即研究所需的样本量。

【例 1-9】以例 1-8 为例，预期两组的变异系数均为 0.5，方差经过对数转换近似相等，FC 为 4，在检验效能 90% 的条件下，试估计各组样本量。

nQuery Advisor 7.0 实现：设定检验水准 $\alpha=0.05$；双侧检验，即 $s=2$；检验效能取 $1-\beta=90\%$ 。依据上述基础数据可知，$FC=4$，$CV=0.5$。

在 nQuery Advisor 7.0 主菜单选择：

Goal：Make Conclusion Using：⊙ Means

Number of Groups：⊙ Two

Analysis Method：⊙ Test

方法框中选择：

Two group t-test for fold change assuming log-normal distribution。

在弹出的样本量计算窗口将各参数键入，如图 1-21 所示，结果为 $n=4$。

Two group t-test for fold change assuming log-normal distribution (equal n's)					
	1	2	3	4	5
Test significance level, α	0.050				
1 or 2 sided test?	2				
Expected Fold-Change, FC	4.000				
Coefficient of variation, CV = σ/u	0.500				
Power (%)	90				
n per group	4				

图 1-21 nQuery Advisor 7.0 关于例 1-9 样本量估计的参数设置与计算结果

SAS 9.2 软件实现：

```
PROC IML;
start MTT0cv(a,s,FC,CV,ratio,power);
error = 0;
if(a > 1 | a < 0) then do;error = 1;print "error"
"Test significance level must be in 0-1";end;
if(s^= 1 & s^= 2) then do;error = 1;print "error"
"s = 1 or 2";end;
if(ratio < 0) then do;error = 1;print "error" "n2/
n1 must be > = 0";end;
if(power > 100 | power < 1) then do;error = 1;
print "error" "Power(%) must be in 1-100";end;
if(error = 1) then stop;
if(error = 0)then do;
es = abs(log(FC))/sqrt(log(1 + CV##2));
n1 = 2;
do until(pw > = power/100);
ncp = sqrt(n1)#es/sqrt(1 + 1/ratio);
```

```
df = 2#(n1-1);
t = tinv(1-a/s,df);
pw = 1-probt(t,df,ncp);
n1 = n1 + 0.01;
end;
n1 = ceil(n1-0.01);
n2 = ceil(ratio#n1);
total = n1 + n2;
print a[label = "Test significance level"]
s[label = "1 or 2 sided test"]
FC[label = "Expected Fold-Change,FC"]
CV[label = "Coefficient of variation,CV"]
power[label = "Power(%)"]
n1[label = "n1"]
n2[label = "n2"];
end;
finish MTT0cv;
run MTT0cv(0.05,2,4,0.5,1,99);
quit;
```

SAS 运行结果:

Test significance level	1 or 2 sided test	Expected Fold-Change,FC	Coefficient of variation,CV	Power(%)	n1	n2
0.05	2	4	0.5	90	4	4

图 1-22　SAS 9.2 关于例 1-9 样本量估计的参数设置与计算结果

1.2.1.4　基于比值的两样本 t 检验

方法：根据 Dilett 等(1991)[8]方法，假设两样本服从对数正态分布，且经对数转换后方差齐。基于比值的两样本 t 检验采用模拟方法，通过反复模拟两独立的服从正态分布的样本，并用两样本 t 检验比较两组均数差异，用有统计差异的次数除以总模拟次数，即所求的检验效能。若要通过 Power 计算样本量 n，可以通过不断尝试样本量 n 的方法找到满足要求的 Power，此时对应的 n 就是要求的样本量。差异存在要满足以下两个条件：

— $P < \alpha$；

— FC 的观察值大于 FCT 的值。

FC 为较大均数和较小均数的比值，FCT 为 FC 允许的最小值，通常有研究的假设来确定，CV 为变异系数，$\sigma = \sqrt{\ln(1+CV^2)}$。

两样本模拟参数的设置，给定初始值 μ_1，根据 FC 和 CV 可分别计算出 $\mu_2, \sigma_1, \sigma_2$。由于采用模拟方法，因此用 SAS 编程得到的结果可能会因为精度和随机种子数等原因与 nQuery 结果有细微差别，但都属正常范围。

【例 1-10】某基因差异性表达的研究中，定义两组的比较只有经检验后差异有统计意义($P < \alpha$)且较大均值与较小均值之比 FC 大于等于 2 时才做进一步研究。假设总体 CV 为 25%，取 $\alpha = 0.05$，双侧检验，预期 FC 值大于 2.1，当检验效能为 80% 时，试估计每组样本量。

nQuery Advisor 7.0 实现：由题意知，$\alpha = 0.05$；$s = 2$；$FCT = 2$；$FC = 2.1$；$CV = 0.25$；$1-\beta = 80\%$。

在 nQuery Advisor 7.0 主菜单选择：

Goal: Make Conclusion Using: ⊙ Means

Number of Groups: ⊙ Two

Analysis Method: ⊙ Test

方法框中选择：

Two group t-test of equal fold change with fold change Threshold。

在弹出的样本量计算窗口将各参数键入，本例模拟次数 10000 次，随机种子数为 22。在 n per group 一栏尝试性填入样本量，最终确定满足检验效能 80% 的最小样本量为 $n = 36$，结果如图 1-23 所示。

nQuery Advisor　[例1-10]

Two group t-test of equal fold change with fold change threshold (equal n's)

	11	12	13	14	15	16	17	18
Test significance level, α	0.050	0.050	0.050	0.050	0.050	0.050	0.050	
1 or 2 sided test?	2	2	2	2	2	2	2	
Fold change threshold, FCT	2.000	2.000	2.000	2.000	2.000	2.000	2.000	
Expected Fold-Change, FC	2.100	2.100	2.100	2.100	2.100	2.100	2.100	
Coefficient of variation, CV = σ / μ	0.250	0.250	0.250	0.250	0.250	0.250	0.250	
Number of simulations	10000	10000	10000	10000	10000	10000	10000	
Random seed for simulations	22	22	22	22	22	22	22	
Probability of detection(%)	90	89	83	81	79	80	80	
n per group	90	80	50	40	35	38	36	

图 1-23　nQuery Advisor 7.0 关于例 1-10 检验效能估计的参数设置与计算结果

SAS 9.2 软件实现：

```
PROC IML;
start MTT0fct(a,s,FCT,FC,CV,n_sti,seed,n);
error = 0;
if(a > 1 | a < 0) then do; error = 1; print "error" "Test significance level must be in 0-1"; end;
if(s^ = 1 & s^ = 2) then do; error = 1; print "error" "s = 1 or 2"; end;
if(n_sti < 0 | ceil(n_sti)^ = n_sti) then do; error = 1; print "error" "The Number of categorys must be positive integer "; end;
if(n < 0 | ceil(n)^ = n) then do; error = 1; print "error" "The n must be positive integer "; end;
if(error = 1) then stop;
if(error = 0) then do;
test = 0;
do i = 1 to n_sti;
mean1 = 1;
mean2 = mean1/FC;
```

(下转第 283 页)

```
gpw = 1-probf( gf, gdf1, gdf2, gncp);
* levels;
lncp = 2#n#M#spher#Vl/inerr##2;
ldf1 = (M-1)#(spher + g1/(n-1));
ldf2 = 2#(n-1)#(M-1)#(spher + g1/(n-1));
lf = finv(1-a, ldf1, ldf2);
lpw = 1-probf(lf, ldf1, ldf2, lncp);
* levels by groups;
glncp = 2#n#M#spher#Vgl/inerr##2;
gldf1 = (M-1)#(spher + g1/(n-1));
gldf2 = 2#(n-1)#(M-1)#(spher + g1/(n-1));
glf = finv(1-a, gldf1, gldf2);
glpw = 1-probf(glf, gldf1, gldf2, glncp);
n = n + 0.01;
end;
gpw = 100 * gpw;
lpw = 100 * lpw;
glpw = 100 * glpw;
n = ceil(n-0.01);
print a[label = "Test significance level"]
M[label = "Number of levels, M"]
Vg[label = "V, between groups"]
Vl[label = "V, between levels"]
Vgl[label = "V, levels by groups"]
bterr[label = "Between-group error term"]
inerr[label = "Within-group error term"]
spher[label = "Measure of 'sphericity'"]
g1[label = "Bias term multiplier"]
gpw[label = "Power, between groups(%)"]
lpw[label = "Power, between levels(%)"]
glpw[label = "Power, levels by groups(%)"]
n[label = "n per group"];
end;
finish MTT3;
run MTT3(0.05, 5, 1.904, 1.582, 0.920, 29.33,
8.13, 0.74, -1.51, 80, 0, 0);
quit;
```

SAS 运行结果：

Test significance level	Number of levels, M	V, between groups	V, between levels	V, levels by groups	Between-group error term	Within-group error term	Measure of 'sphericity'
0.05	5	1.904	1.582	0.92	29.33	8.13	0.74

Bias term multiplier	Power, between groups(%)	Power, between levels(%)	Power, levels by groups(%)	n per group
-1.51	80.000732	99.999996	99.976959	356

图1-34 SAS 9.2 关于例1-13样本量估计的参数设置与计算结果

（上接第278页）

```
sd1 = mean1#CV;
sd2 = mean2#CV;
group1 = normal(repeat(seed, n, 1))#sd1 +
mean1;
group2 = normal(repeat(seed, n, 1))#sd2 +
mean2;
FC_ = group1[:]/group2[:];
CV_ = 1/2#( sqrt(sum((group1-group1[:])##
2)/(n-1))/group1[:] + sqrt(sum((group2-group2
[:])##2)/(n-1))/group2[:]);
if FC_ > 0 then do;
t = sqrt(n)#abs(log(FC_))/sqrt(2#log(1 + CV_##2));
df = 2#(n-1);
t1 = tinv(1-a/s, df);
t2 = tinv(a/s, df);
if( FC_ > = FCT &(t > = t1 | t < = t2)) then test
= test + 1;
end;
end;
power = floor(test/n_sti#100);
print a[label = "Test significance level"]
s[label = "1 or 2 sided test"]
FCT[label = "Fold-change threshold, FCV"]
FC[label = "Expected fold-change, FC"]
seed[label = "Random seed for simulations"]
n_sti[label = "Number of simulations"]
power[label = "Power(%)"]
n[label = "n"];
end;
finish MTT0fct;
run MTT0fct(0.05, 2, 2, 2.1, 0.25, 10000, 202, 36);
quit;
```

SAS 运行结果：

Test significance level	1 or 2 sided test	Fold-change threshold, FCV	Expected fold-change, FC	Random seed for simulations	Number of simulations	Power (%)	n
0.05	2	2	2.1	202	10000	80	36

图1-24 SAS 9.2 关于例1-10检验效能估计的参数设置与计算结果