

## 样本量估计及其在 nQuery 和 SAS 软件上的实现\* ——率的比较(二)

张惠凤<sup>1</sup> 揭著业<sup>1</sup> 唐欣然<sup>1</sup> 陈平雁<sup>2△</sup>

### 2.1.1.5 多分类样本率(单组构成比)的 $\chi^2$ 检验

方法:本文所述的多分类样本率的  $\chi^2$  检验实质上是指样本构成比与总体构成比的  $\chi^2$  检验。Lachin (1977)<sup>[4]</sup> 提出的此类资料的样本量估计是建立在非中心的  $\chi^2$  分布基础上的。其检验效能的计算公式为:

$$1 - \beta = 1 - \chi^2(\chi^2_{1-\alpha, (c-1)}, c-1, n\Delta^2) \quad (2-11)$$

式中,  $c$  为分类数;  $\Delta^2$  为效应量,  $n\Delta^2$  为非中心参数;  $\chi^2(x, df, np)$  是非中心  $\chi^2$  分布的累积分布函数。  $\Delta^2$  的计算公式如下,

$$\Delta^2 = \sum_{i=1}^c \frac{(\pi_{0i} - \pi_{1i})^2}{\pi_{0i}} \quad (2-12)$$

式中,  $\pi_{0i}$  为原假设下(总体)第  $i$  类的构成比,  $\pi_{1i}$  为备择假设下(样本)第  $i$  类的构成比;值得注意的是,这里假设每类的样本量  $n$  相等。样本量可以通过逼近的方法求得。

【例 2-5】某研究欲比较某医院感染病例三类病原菌( $G^+$  菌、 $G^-$  菌和真菌)的比例与一般医院有无差异,根据已有的研究报道,一般医院感染病例中三类病原菌的分布为  $G^+$  菌(35%)、 $G^-$  菌(50%)、真菌(15%),假设通过预调查获得某医院感染病例的三类病原菌分布为  $G^+$  菌(31.2%)、 $G^-$  菌(57.6%)、真菌

(11.2%),试估计检验效能为 80% 所需的样本量。

nQuery Advisor7.0 实现:设定检验水准  $\alpha = 0.05$ ; 检验效能取  $1 - \beta = 80\%$ 。已知  $c = 3$ , 总体构成比为 0.35、0.50、0.15;预期的总体构成比为 0.312、0.576、0.112。在 nQuery Advisor7.0 主菜单选择:

Goal: Make Conclusion Using:  Proportions

Number of Groups:  One

Analysis Method:  Test

方法框中选择: Chi-square test of specified proportions in C categories。

注意,这里首先应根据总体和样本构成比对效应量  $\Delta^2$  进行估计,在菜单栏中选择:

Assistants:  Compute Effect Size

弹出辅助计算窗口。依据上述基础数据,总体构成比为 0.350、0.500、0.150,样本构成比为 0.312、0.576、0.112,在弹出的计算窗口将各参数值键入,如图 2-10 所示,结果为  $\Delta^2 = 0.025$ 。

点击 Transfer 后,将所计算参数传输至样本量估计窗口,如图 2-11 所示,结果为  $n = 381$ 。

Compute	Transfer	Close	Restore	Clear	Cut	Copy	Paste	Print
			Cat. 1	Cat. 2	Cat. 3	$\sum_i \pi_{1i}$		
			Null ( $H_0$ ) proportions	0.350	0.500	0.150	1.000	
			Alternative ( $H_1$ ) proportions	0.312	0.576	0.112	1.000	
			$\Delta^2 = \sum\{(\pi_{0i} - \pi_{1i})^2 / \pi_{0i}\}$	0.025				

图 2-10 nQuery Advisor7.0 关于例 2-5 样本量估计中效应量的计算

Chi-square test of specified proportions in C categories					
	1	2	3	4	5
Test significance level, $\alpha$	0.050				
Number of categories, C	3				
Effect size, $\Delta^2 = \sum(\pi_{0i} - \pi_{1i})^2 / \pi_{0i}$	0.0253				
Power (%)	80				
n	381				

图 2-11 nQuery Advisor7.0 关于例 2-5 样本量估算的参数设置与计算结果

\*:教育部重点课题(DIA070113)资助

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

SAS9.2 软件实现:

```
% macro POT2(alpha, c, d, power);
```

```
data POT2;
```

```
alpha = &alpha;
```

```

c = &c;
d = &d;
power = &power;
n = 0;
do until ((power2 - power) >= 10 * * (-7));
power2 = 1 - cdf('CHISQUARE', cinv(1 - alpha, c - 1), c - 1, n * d);
n = n + 1;
end;
n = n - 1;
keep alpha c d power n;
run;

```

```

proc print data = POT2 label;
var alpha c d power n;
label alpha = 'Test significance level'
c = 'Numbers of categories'
d = 'effect size'
power = 'Power'
n = 'n';
quit;
% mend POT2;
% POT2(0.05, 3, 0.0253, 0.80);
SAS 运行结果:

```

Test significance level	Numbers of categories	effect size	Power	n
0.05	3	0.0253	0.8	381

图 2-12 SAS9.2 关于例 2-5 样本量估计的参数设置与计算结果

2.1.2 等效性/非劣性检验

2.1.2.1 配对设计一致率的等效性检验

方法: Tango (1998)<sup>[5]</sup> 根据正态近似法提出该检验的样本量估计方法, 公式如下:

$$n = \frac{\eta [z_{1-\alpha/s} + z_{1-\beta}]^2}{(\Delta_0 - \Delta_1)^2} \quad (2-13)$$

其中,

$$\eta = \pi_{01} + \pi_{10}, \Delta_0 = \pi_T - \pi_S \quad (2-14)$$

式中,  $s$  在 nQuery Advisor 7.0 软件中默认为 1, 即单侧检验;  $\eta$  为结果不一致的对子数占总对子数的比例, 其含义见表 2-1;  $\Delta_0$  为等效性界值, 即应用方面可以接受的试验组与对照组总体阳性率的差值;  $\Delta_1$  为预期两个总体阳性率之差, 等效性检验通常取  $\Delta_1 = 0$ , 但  $\Delta_1$  也可以不取 0; 相应的检验效能由上述公式反推求得。

【例 2-6】某临床试验欲采用配对设计, 证实一种新的较廉价的某疾病的诊断方法其灵敏度与现有标准诊断方法等效。已知现有标准诊断方法的灵敏度为 85%, 假设新诊断方法的灵敏度预期为 85%, 两种诊断方法都为阳性结果的比例占 80%, 临床应用的等效界值为 5%, 那么当检验效能为 85% 时所需的样本量是多少?

nQuery Advisor 7.0 实现: 设定单侧检验水准  $\alpha = 0.025$ ; 检验效能取  $1 - \beta = 85\%$ 。依题意可知,  $\Delta_0 = \Delta_T - \Delta_S = -0.05$ ,  $\Delta_1 = 0.85 - 0.85 = 0$ , 两种诊断结果均为阳性的比例为 80%, 可得到表 2-1 数据, 由此推算出结果不一致的对子占总对子数的比例为  $\eta = 0.05 + 0.05 = 0.10$ 。

在 nQuery Advisor 7.0 主菜单选择:

Goal: Make Conclusion Using:  Proportions

Number of Groups:  One

Analysis Method:  Equivalence

表 2-1 例 2-6 所形成的数据

新诊断方法	标准诊断方法	
	+	-
+	$\pi_{11}$ (80%)	$\pi_{10}$ (5%)
-	$\pi_{01}$ (5%)	$\pi_{00}$ (10%)

方法框中选择: Paired responses: Equivalence of proportions。

在弹出的样本量估计窗口将各参数值键入, 如图 2-13 所示, 结果为  $n = 360$ 。需要指出, 若  $\Delta_1$  取 0, 则  $\Delta_0$  的符号可以忽略; 否则  $\Delta_0$  的符号就非常重要了, 一定要定义好差值的方向。

SAS9.2 软件实现:

```

% macro POE0(a, d0, d1, e, power);
data POE0;
a = &a;
d0 = &d0;
d1 = &d1;
e = &e;
s = 1;
power = &power;
n = e * (probit((1-a)/s) + probit(power)) * *
2 / (d0 - d1) * * 2;
n = ceil(n);
run;
proc print data = POE0 label;
var a d0 d1 e power n;
label a = 'Test significance level'
d0 = 'equivalence limit difference'
d1 = 'expected difference'

```

e = 'proportion discordant'  
power = 'Power'  
n = 'n';  
quit;

% mend POE0;  
% POE0(0.05, -0.05, 0, 0.1, 0.85);  
SAS 运行结果:

	1	2	3	4
Test significance level, $\alpha$ (one-sided)	0.025			
Equivalence limit difference, $\Delta_0$	-0.050			
Expected difference, $\Delta_1$	0.000			
Proportion discordant, $\eta = \pi_{10} + \pi_{01}$	0.100			
Power (%)	85			
n (number of pairs)	360			

图 2-13 nQuery Advisor7.0 关于例 2-6 样本量估计的参数设置与计算结果

Test significance level	equivalence limit difference	expected difference	proportion discordant	Power	n
0.025	-0.05	0	0.1	0.85	360

图 2-14 SAS9.2 关于例 2-6 样本量估计的参数设置与计算结果

2.1.2.2 配对设计基于非劣效下线的检验(试验组-对照组)

方法:Newcombe(1988)<sup>[6]</sup>提出的配对设计的二分类数据的等效性或非劣性验证的样本量估计是基于可信区间的方法,通过模拟实现的。首先模拟产生四格表资料,然后使用 Newcombe-Wilson score 的连续校正方法计算每一次模拟数据的可信区间,如果该区间落在事先确定的等效或非劣效区间内,则记为成功,总的成功次数除以总的模拟次数所得的就是所需要的检验效能。非劣效下限检验用于非劣效验证中试验组的阳性率虽低于对照组但其差异不大于某一界值的设计。

四格表的产生及资料形式见表 2-2,这里  $\pi_{11}$  表示对照组和实验组均为阳性事件的概率,频数为  $e$ ,  $\pi_{10}$  表示对照组阳性而实验组阴性的概率,频数为  $f$ ,  $\pi_{01}$  表示对照组阴性而实验组阳性的概率,频数为  $g$ ,  $\pi_{00}$  表示对照组和实验组均为阴性事件的概率,频数为  $h$ 。首先,使用二项分布随机数产生对照组成功的一系列  $n_1$ ,参数为  $n$  和  $\pi_{11} + \pi_{10}$ ,对照组不成功为  $n - n_1$ ;再使用二项随机数产生  $e$ ,参数为先前产生的  $n_1$  和  $\pi_{11}/(\pi_{11} + \pi_{10})$ ;  $f = n_1 - e$ ;其他的类似。

表 2-2 四格表的产生及资料形式\*

对照组	实验组		合计
	阳性	阴性	
阳性	$e(\pi_{11})$	$f(\pi_{10})$	$n_1$
阴性	$g(\pi_{01})$	$h(\pi_{00})$	$n - n_1$
合计	$m$	$n - m$	$n$

\* 括弧外为频数,括弧内为概率

Newcombe-Wilson Score 非连续校正可信区间的计算公式:区间从  $L = \hat{\theta} - \delta$  到  $U = \hat{\theta} + \varepsilon$ ,其中,

$$\hat{\theta} = |\pi_{10} - \pi_{01}|, \delta = \sqrt{\{dl_2^2 - 2\phi dl_2 du_3 + du_3^2\}} \quad (2-15)$$

$$dl_2 = (e + f)/n - l_2, du_2 = u_2 - (e + f)/n \quad (2-16)$$

$l_2$  和  $u_2$  是关于  $\zeta$  的方程  $|\zeta - (e + f)/n| = z_{1-\alpha/s} \sqrt{\zeta(1-\zeta)/n}$  的根。类似的,

$$dl_3 = (e + g)/n - l_3, du_3 = u_3 - (e + g)/n \quad (2-17)$$

$l_3$  和  $u_3$  是关于  $\zeta$  的方程  $|\zeta - (e + g)/n| = z_{1-\alpha/s} \sqrt{\zeta(1-\zeta)/n}$  的根。同样,

$$\hat{\theta} = (eh - fg) / \sqrt{\{(e + f)(g + h)(e + g)(f + h)\}} \quad (2-18)$$

Newcombe-Wilson Score 连续性校正可信区间:在非连续性校正可信区间基础上对  $\hat{\theta}$  进行校正,如果  $eh > fg$ ,  $\hat{\theta}$  的分子等于  $\max(eh - fg - n/2, 0)$ 。

配对设计基于非劣效下线的检验效能:对于每个模拟出来的四格表,运用 Newcombe-Wilson Score 方法求出的可信区间下限大于事先设定的区间下限,就算成功 1 次,总的成功次数除以总的模拟次数所得的就是检验效能。

【例 2-7】某临床试验欲采用配对设计非劣效验证。已知现有标准诊断方法的准确度为 85%,假设新诊断方法的准确度预期为 85%,两种诊断方法不一致的对子占总对子的比例  $\eta$  为 0.1,临床认可的非劣效界值为 8%,那么当检验效能分别为 80% 和 90% 时所需的样本量是多少?

nQuery Advisor7.0 实现:非劣效验证采用单侧检验,设定检验水准  $\alpha = 0.025$ ;检验效能  $1 - \beta = 90\%$ 。非劣效界值  $\Delta_0$  为  $-0.08$ ,预期的率差  $\Delta_1$  为 0。种子设为 1000,模拟次数为 10000 次。在 nQuery Advisor7.0 主菜单选择:

Goal: Make Conclusion Using:  Proportions  
Number of Groups:  One group  
Analysis Method:  Equivalence

方法框中选择: Paired responses; Equivalence of Proportions。

2-15 所示,当检验效能为 80% 时样本量为 140;检验效能为 90% 时样本量为 182。

在弹出的样本量估计窗口将各参数值键入,如图

	1	2	3	4	5
Confidence level, 1- $\alpha$ (one-sided)	0.975	0.975			
Expected difference, $\pi_T - \pi_S, \Delta_1$	0.000	0.000			
Proportion discordant, $\eta = \pi_{10} + \pi_{01}$	0.100	0.100			
Proportion both yes, $\pi_{11}$	0.800	0.800			
Lower limit for $\pi_T - \pi_S, LL$	-0.080	-0.080			
Number of simulations	10000	10000			
Random seed for simulations	1000	1000			
Power (%)	80	90			
n (number of pairs)	140	182			

图 2-15 nQuery Advisor7.0 关于例 2-7 样本量估计的参数设置与计算结果

SAS9.2 软件实现:

```
% macro POE1_1(st, seed, n, de, h, p1, side, alpha,
1);
data POE1_1;
st = &st;
seed = &seed;
n = &n;
p1 = &p1;
de = &de;
h = &h;
alpha = &alpha;
side = &side;
za = probit(1-alpha/side);
l = &l;
i = 1;
pr = 0;
p2 = (de + h)/2;
p3 = (h-de)/2;
p4 = 1-p2-p3-p1;
do until(i > st);
fr1 = ranbin(seed, n, p1 + p2);
e = ranbin(seed, fr1, p1/(p1 + p2));
f = fr1 - e;
fr2 = n - fr1;
g = ranbin(seed, fr2, p3/(p3 + p4));
h = fr2 - g;
p1a = (e + f)/n;
p2a = (e + g)/n;
l2 = (2 * n * p1a + za * * 2 - za * sqrt(za * * 2 + 4 * n * p1a * (1-p1a)))/(n + za * * 2)/2;
u2 = (2 * n * p1a + za * * 2 + za * sqrt(za * * 2 + 4 * n * p1a * (1-p1a)))/(n + za * * 2)/2;
l3 = (2 * n * p2a + za * * 2 - za * sqrt(za * * 2 + 4 * n * p2a * (1-p2a)))/(n + za * * 2)/2;
u3 = (2 * n * p2a + za * * 2 + za * sqrt(za * * 2 +
```

```
4 * n * p2a * (1-p2a)))/(n + za * * 2)/2;
dl2 = (e + f)/n - l2;
du2 = u2 - (e + f)/n;
dl3 = (e + g)/n - l3;
du3 = u3 - (e + g)/n;
if e * h > f * g then fz = max(e * h - f * g - n/2, 0);
else fz = e * h - f * g;
o = fz/sqrt((e + f) * (g + h) * (e + g) * (f + h));
ld = sqrt(dl2 * * 2 - 2 * o * dl2 * du3 + du3 * * 2);
rd = sqrt(du2 * * 2 - 2 * o * du2 * dl3 + dl3 * * 2);
if (p1a - p2a - ld > = 1) then pr = pr + 1;
i = i + 1;
end;
power = pr/st;
h = &h;
run;
proc print data = POE1_1 label;
var alpha side p1 de h l power n;
label alpha = 'Test significance level'
side = '1 or 2 side' p1 = 'proportion1'
de = 'expected difference' h = 'proportion discordant'
l = 'low limite' power = 'Power' n = 'n';
quit;
% mend POE1_1;
% POE1_1(st = 10000, seed = 1000, n = 137, de = 0,
h = 0.1, p1 = 0.8, side = 2, alpha = 0.05, l = -0.08);
% POE1_1(st = 10000, seed = 1000, n = 182, de = 0,
h = 0.1, p1 = 0.8, side = 2, alpha = 0.05, l = -0.08);
SAS 运行结果:
SAS9.2 模拟结果和 nQuery Advisor7.0 稍有偏差,这可能是由于两种软件在模拟过程随机数发生机制不同所导致。
```

Test significance level	number of simulations	random seed for simulation	1 or 2 side	proportion1	expected difference	proportion discordant	low limite	Power	n
0.025	10000	1000	1	0.8	0	0.1	-0.08	0.8003	137

  

Test significance level	number of simulations	random seed for simulation	1 or 2 side	proportion1	expected difference	proportion discordant	low limite	Power	n
0.025	10000	1000	1	0.8	0	0.1	-0.08	0.9024	182

图 2-16 SAS9.2 关于例 2-7 样本量估计的参数设置与计算结果

2.1.2.3 配对设计基于非劣效上限的检验(试验组-对照组)

方法:将上述的配对设计基于非劣效下限检验的估计方法由下限改为上限,就是配对设计基于非劣效上限检验的样本量估计方法,这里就不赘述。

2.1.2.4 配对设计双单侧等效性检验(试验组-对照组)

方法:将配对设计基于非劣效下限检验和上限检验的样本量估计结合在一起就是所谓的双单侧等效性检验的样本量估计。概括而言,对于每个模拟出来的四格表,如果运用 Newcombe-Wilson Score 方法求出的置信区间在我们事先设定的区间内,就算成功 1 次,总的成功次数除以总的模拟次数得到的就是检验效能。

【例 2-8】以例 2-6 为例,应用 Newcombe-Wilson Score 方法估计样本量。

nQuery Advisor7.0 实现:等效性验证采用双单侧检验,设定每个单侧的检验水准  $\alpha = 0.025$ ,即对应的

可信区间范围为 97.5%;检验效能  $1-\beta = 85%$ ,非劣效界值的下限为-5%,上限为 5%。

在 nQuery Advisor7.0 主菜单选择:

Goal:Make Conclusion Using:⊙ Proportions

Number of Groups:⊙ One group

Analysis Method:⊙ Equivalence

方法框中选择:Paired responses; Equivalence of Proportions (using Confidence Interval)。

在弹出的样本量估计窗口将各参数值键入,如图 2-17 所示,求得样本量为 569 例。需要指出,例 2-6 应用正态逼近方法求得的样本量为 360 例,两种方法所得结果的差距较大,我们将对此问题进行深入研究,尽早给出这一问题的研究结论。

SAS9.2 软件实现:

```
% macro POE1_lu ( st, seed, n, de, h, p1, side, alpha, l, r );
data POE1_lu;
st = &st;
```

	1	2	3	4	5
Confidence level, 1- $\alpha$	0.975				
Expected difference, $\pi_T - \pi_S, \Delta_1$	0.000				
Proportion discordant, $\eta = \pi_{10} + \pi_{01}$	0.100				
Proportion both yes, $\pi_{11}$	0.800				
Lower limit for $\pi_T - \pi_S, LL$	-0.050				
Upper limit for $\pi_T - \pi_S, UL$	0.050				
Number of simulations	10000				
Random seed for simulations	1000				
Power (%)	85				
n (number of pairs)	569				

图 2-17 nQuery Advisor7.0 关于例 2-8 样本量估计的参数设置与计算结果

```
seed = &seed;
n = &n;
p1 = &p1;
de = &de;
h = &h;
alpha = &alpha;
side = &side;
za = probit( 1-alpha/side );
l = &l;
r = &r;
i = 1;
pr = 0;
```

```
p2 = ( de + h ) / 2;
p3 = ( h - de ) / 2;
p4 = 1 - p2 - p3 - p1;
do until ( i > st );
fr1 = ranbin ( seed, n, p1 + p2 );
e = ranbin ( seed, fr1, p1 / ( p1 + p2 ) );
f = fr1 - e;
fr2 = n - fr1; g = ranbin ( seed, fr2, p3 / ( p3 + p4 ) );
h = fr2 - g;
p1a = ( e + f ) / n;
p2a = ( e + g ) / n;
```

(下转第 767 页)

显著的;如果个体在初次干预后就不吸烟了,那么再进行干预也就失去了意义,当然分组效果是不显著的。但是如果个体在初次干预后仍然吸烟,则时刻 2 是否吸烟受到分组的影响,而且处理组的吸烟率高于对照组,这说明初次干预没有戒掉烟瘾的吸烟者,对于后面的多次干预是没有显著效果的,从长期来看,反倒是对照组中依靠自我控制来戒烟的效果更明显。

综合上述分析我们得到的结论是:

第一,缺失与分组有关,位于对照组的个体缺失的比例更大,缺失个体吸烟的可能性低于可观测到个体,这说明多数缺失的个体已戒除烟瘾,不需要继续参加戒烟训练,同时说明文献[1]中 missing = smoking 的补值方法是不合理的;

第二,在时刻 1 干预效果显著,吸烟与分组有关,处理组的戒烟率高于对照组;如果不考虑初次干预后个体吸烟的状态,则无论是对有缺失数据的情况,还是对于通过 EM 算法补值后得到完全数据的情况,戒烟效果与个体所在组无关;

第三,如果初次干预后个体不吸烟了,则在 24 个月,个体是否吸烟也与所在组无关;

第四,如果初次干预后个体仍然吸烟,则在 24 个月,个体是否吸烟与所在组有关,处理组的戒烟方法没有明显的效果。

回到试验背景,我们看到初次干预结果是非常重要的,通过自我控制来戒烟的方式从长期来看是更加有效的。这也为今后的戒烟训练工作提供了一个理论指导。

本文的结论更符合实际背景,这在一定程度上说明本文所建立的图模型是合理的。与文献[1]相比,本文的补值方法可信度更高,可用于分析其他含有缺失数据的两值数据,本文得到的结论对戒烟训练工作更有参考价值。

参 考 文 献

1. Hedeker D, Robin J, Demirtas H. Analysis of binary outcomes with missing data: missing = smoking, last observation carried forward, and a little multiple imputation. Society for the Study of Addiction, 2007, 102: 1564-1573.
2. Ma WQ, Geng Z, Hu YH. Identification of graphical models for non-ignorable nonresponse of binary outcome in longitudinal studies. Journal of Multivariate Analysis, 2003, 87: 24-45.
3. Little RJA, Rubin DR. Statistical analysis with missing data. New York: Wiley, 1987.
4. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society, Series B, 1977, 39: 1-38.
5. 赵志文, 王思洋, 王瑞庭, 等. 定时截尾下具有部分缺失数据两个指数总体参数估计与检验. 吉林大学学报(理学版), 2009, 47(1): 26-30.

(上接第 763 页)

```

l2 = (2 * n * p1a + za * * 2 - za * sqrt( za * * 2 + 4
* n * p1a * (1-p1a) )) / (n + za * * 2) / 2;
u2 = (2 * n * p1a + za * * 2 + za * sqrt( za * * 2 +
4 * n * p1a * (1-p1a) )) / (n + za * * 2) / 2;
l3 = (2 * n * p2a + za * * 2 - za * sqrt( za * * 2 + 4
* n * p2a * (1-p2a) )) / (n + za * * 2) / 2;
u3 = (2 * n * p2a + za * * 2 + za * sqrt( za * * 2 +
4 * n * p2a * (1-p2a) )) / (n + za * * 2) / 2;
dl2 = (e + f) / n - l2;
du2 = u2 - (e + f) / n;
dl3 = (e + g) / n - l3;
du3 = u3 - (e + g) / n;
if e * h > f * g then fz = max(e * h - f * g - n / 2, 0);
else fz = e * h - f * g;
o = fz / sqrt( (e + f) * (g + h) * (e + g) * (f +
h) );
ld = sqrt( dl2 * * 2 - 2 * o * dl2 * du3 + du3 * * 2 );
rd = sqrt( du2 * * 2 - 2 * o * du2 * dl3 + dl3 * * 2 );
if (p1a - p2a - ld > = 1) && (p1a - p2a + rd < = r)

```

```

then
pr = pr + 1;
i = i + 1;
end;
power = pr / st;
h = &h;
run;
proc print data = POE1_lu label;
var alpha side p1 de h l r power n;
label alpha = 'Test significance level'
side = '1 or 2 side' p1 = 'proportion1'
de = 'expected difference' h = 'proportion discordant'
l = 'lower limite' r = 'upper limite' power = 'Power'
n = 'n'; quit;
%mend POE1_lu;
%POE1_lu( st = 10000, seed = 1000, n = 569, de =
0, h = 0.1, p1 = 0.8, side = 2, alpha = 0.025, l = -0.05, r
= 0.05 );
SAS 运行结果:

```

Test significance level	1 or 2 side	proportion1	expected difference	proportion discordant	lower limite	upper limite	Power	n
0.025	2	0.8	0	0.1	-0.05	0.05	0.8515	569

图 2-18 SAS9.2 关于例 2-8 样本量估计的参数设置与计算结果