

# 基于 SAS 软件实现样本含量估计及应用

陶丽新<sup>1,2</sup>, 刘一松<sup>3</sup>, 胡良平<sup>2,4\*</sup>

(1. 首都医科大学公共卫生学院, 北京 100069;

2. 世界中医药联合会临床科研统计学专业委员会, 北京 100029;

3. 北京医普科诺科技有限公司, 北京 100190;

4. 军事医学科学院生物医学统计学咨询中心, 北京 100850

\* 通信作者: 胡良平; E-mail: lphu812@sina.com)

**【摘要】** 本文的目的是使读者能方便快捷地运用 SAS 软件中的 POWER 过程和 GLMPOWER 过程实现样本含量估计。在不同的场合下估计样本含量需要提供不同的前提条件, 即使仅限于假设检验时估计样本含量, 也必须进一步弄清对应的设计类型、结果变量的性质、某些先验知识和对结果精确度的要求。本文通过一些实例, 介绍了估计样本含量与检验效能的 SAS 实现方法。读者只需要修改本文中所呈现的 SAS 程序中的少量参数, 就可方便地使用 SAS 软件实现样本含量与检验效能估计。事实说明, 尽管 SAS 软件非常难学难用, 但借助现成的 SAS 程序, 读者可以轻松自如地解决很多与统计分析有关的具体问题。

**【关键词】** SAS 软件; 样本含量; 检验效能; 均值; 标准差; 率

中图分类号: R195.1

文献标识码: A

doi: 10.11886/j.issn.1007-3256.2016.05.004

## Sample size estimation and its application by using SAS software

TAO Li-Xin<sup>1,2</sup>, LIU Yi-song<sup>3</sup>, HU Liang-ping<sup>2,4\*</sup>

(1. Public Health Academy, Capital University of Medical Sciences, Beijing 100069, China;

2. Specialty Committee of Clinical Scientific Research Statistics of World Federation of Chinese Medicine Societies, Beijing 100029, China;

3. Beijing Improve-Quality Tech Co., Ltd., Beijing 100190, China;

4. Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China

\* Corresponding author: HU Liang-ping, E-mail: lphu812@sina.com)

**【Abstract】** The paper aims to help the readers to estimate the sample size by using the POWER and GLMPOWER procedures in SAS software conveniently and easily. It is necessary for a user to supply the prerequisites of sample size estimation under the different situations, even if limited under the case of the hypothesis testing. Usually the premise conditions of sample size estimation, which should be given, are as follows: the design type, the characters of outcomes, the prior knowledge, and the accuracy of the results. The methods of estimating the sample size and power by SAS were introduced through several real examples in this article. It is convenient for the readers to realize the estimation of sample size and power by using SAS. The only thing for them to do is to modify a few parameters in the given SAS programs. Although SAS software is very difficult for people to learn and use, as a matter of fact, the users can solve the concrete problems concerned with statistical analysis by means of revisions of ready-made SAS programs easily.

**【Key words】** SAS software; Sample size; Power; Mean value; Standard deviation; Rate

### 1 估计样本含量的必要性与前提条件

在开展试验性科学研究中, 往往需要在相同试验条件下做多次重复试验, 以便使随机变量的变化规律能充分显露出来。文献[1]就“重复原则与样本含量”、“重复的三层含义”、“估计样本含量和检验效能的意义”、“估计样本含量需要的前提条件”等内容, 作了较为详细的介绍。本文将直接运用 SAS 软件中的 POWER 过程和 GLMPOWER 过程估计样本含量和检验效能。

### 2 定量资料假设检验时样本含量估计<sup>[2-5]</sup>

**【例 1】**某研究者欲比较两种物理疗法增加肌肉弹性的效果。希望采用  $t$  检验比较两总体均值之间差别并希望得出具有统计学意义的检验结果, 设检

验水准  $\alpha$  为 0.05, 检验效能  $1-\beta$  为 0.9, 假定的均值和标准差的数值(在实践中, 这些数据应基于文献资料或预实验的结果) 将列在下面。拟回答的问题是: 至少需要检测多少例受试对象?

标准治疗方法的肌肉弹性均值为 13, 如果新疗法的肌肉弹性被定位 14.0、14.5 和 15.0 三种情况, 且假定分别有两种可能的标准差, 即  $\sigma = 1.2$  与  $\sigma = 1.7$ ; 假定研究者还希望两组的样本含量具有下列三种比例关系, 即: ①每个组的样本含量相同(1:1); ②新疗法的样本含量是标准疗法的两倍(1:2); ③新疗法的样本含量是标准疗法的三倍(1:3)。

**【分析与解答】**根据前面给出的条件可知, 均值有 3 种可能的取值, 标准差有两种可能的取值, 两组样本含量的比例有三种可能的取值, 故这样一共有

3 × 2 × 3 = 18 种情况。也就是说,要在 18 种条件下估计所需要的样本含量。

使用下面的 SAS 程序,可以获得与上面 18 种条件对应的样本含量估计结果:

```
proc power;
  twosamplemeans
    groupmeans = (13 14) (13 14.5) (13 15)
    stddev = 1.2 1.7
    groupweights = 1 | 1 2 3
    power = 0.9
    ntotal = .;
run;
```

**【程序说明】**在 power 过程中使用 twosamplemeans 语句来决定 18 种情况下为了达到 90% 的检验效能所需要的样本含量。将选项 “ntotal = .” 指定样本含量为结果变量;用 “groupmeans = ” 选项指定假设的均值,对每种情况的均值使用匹配标记,即用圆括号隔开不同的均值组合条件;使用 “stddev = ” 选项为每种情况指定标准差;使用 “groupweights = ” 选项指定权重比例,可以再使用匹配标记,即使用 “|” 符号隔开不同的取值。

**【输出结果】**

Computed N Total						
Index	Mean1	Mean2	Std Dev	Weight2	Actual Power	N Total
1	13	14.0	1.2	1	0.907	64
2	13	14.0	1.2	2	0.908	72
3	13	14.0	1.2	3	0.905	84
4	13	14.0	1.7	1	0.901	124
5	13	14.0	1.7	2	0.905	141
6	13	14.0	1.7	3	0.900	164
7	13	14.5	1.2	1	0.910	30
8	13	14.5	1.2	2	0.906	33
9	13	14.5	1.2	3	0.916	40
10	13	14.5	1.7	1	0.900	56
11	13	14.5	1.7	2	0.901	63
12	13	14.5	1.7	3	0.908	76
13	13	15.0	1.2	1	0.913	18
14	13	15.0	1.2	2	0.927	21
15	13	15.0	1.2	3	0.922	24
16	13	15.0	1.7	1	0.914	34
17	13	15.0	1.7	2	0.921	39
18	13	15.0	1.7	3	0.910	44

**【结果解读】**最好的设计类型是(均值差最大为 2,小的标准差为 1.2 和平衡设计),达到至少 0.9 的检验效能需要的样本含量为 18(  $n_1 = n_2 = 9$  );最差的设计为(小的均值差为 1,大的标准差为 1.7,1:3 的非平衡设计),达到至少 0.9 的检验效能需要样本含量为 164(  $n_1 = 41, n_2 = 123$  )。在“固定设计中”指定的检验效能为 0.9,在“计算的样本含量”表格中输出根据样本比例调整的实际的检验效能(见上面输出结果中倒数第 2 列)。

值得一提的是,在上面的 SAS 程序中,若给定总样本含量,而将 power 设置为“缺省值”就是在各

种不同条件下,估计检验效能。如将上面的 SAS 程序修改如下:

```
proc power;
  twosamplemeans
    groupmeans = (13 14) (13 14.5) (13 15)
    stddev = 1.2 1.7
    groupweights = 1 | 1 2 3
    power = .
    ntotal = 100;
run;
```

读者可以自己运行上面这段 SAS 程序,会得到各种条件下的检验效能估计值,因篇幅所限,此输出

结果从略。在下面的其他 SAS 程序中,也可进行同样的修改。

**【例 2】**研究轻度和重度再生障碍性贫血患者血清中可溶性 CD<sub>8</sub> 抗原水平(U/mL)与正常人之间的差别,以反映患者免疫状态紊乱而导致造血功能障碍的程度,共设置正常组、轻度组和重度组 3 个实验组。根据以往研究得知 3 种人群的 CD<sub>8</sub> 抗原水平( $\bar{x} \pm s$ )分别为:正常组(290 ± 174) 轻度组(658 ± 155), 重度组(763 ± 127)。假定本研究各组的样本含量相等,设检验水准  $\alpha$  为 0.05 检验效能  $1 - \beta$  为 0.8, 试估计所需的样本含量。

**【分析与解答】**这是一个结果变量为定量变量的单因素三水平设计试验研究的样本含量估计问题。所需的 SAS 程序如下。

```
ods html;
proc power;
onewayanova groupmeans =290 | 658 | 763 stddev = 174 155 127
groupweights = (1 1 1) alpha = 0.05 Ntotal =. power =0.8;
run; quit;
ods html close;
```

**【程序说明】**选项“groupmeans =290 | 658 | 763”定义各水平组指标总体均值的估计值;选项“stddev = 174 155 127”定义几种总体标准差的估计值;选项“groupweights = (1 1 1)”定义各水平组样本含量的分配比例。对于类似的问题,上述程序中的数字需要根据用户的实际情况进行修改。

#### 【输出结果】

Index	Std Dev	Actual Power	Computed N Total
1	174	0.866	12
2	155	0.933	12
3	127	0.911	9

**【结果解读】**分析结果表明:本研究各类人群只需随机抽取 4 例 3 类人群共 12 例即可满足要求。

**【例 3】**用三种方法治疗脑卒中抑郁患者,观察神经功能康复状况。估计治疗后三种方法斯堪地那维亚卒中量表(Scemelinavian Stroke Scale, SSS)评分均数分别为 11.0、23.0、9.0 标准差分别为 3、3、2。希望各组样本含量的比例或权重(本例取 1:1:1)、各组的平均标准差(保守的做法,可采用所有组的最大标准差)、允许犯第 I 类(或假阳性)错误的概率  $\alpha$ (通常取其值为 0.05)、给定检验效能  $1 - \beta$  的值

[相当于给定允许犯第 II 类(或假阴性)错误的概率  $\beta$ (通常取其值为 0.2 或 0.1) 本例取  $1 - \beta = 0.9$ ]。如果要得到三组之间的差别以及两两比较的结果都有统计学意义的结论,问每组各需要多少例患者?

**【分析与解答】**这是做单因素三水平设计一元定量资料均值比较及两两比较之前进行样本含量估计的问题。所需的 SAS 程序如下。

```
data d11_20;
input method $ score CellWgt;
datalines;
A 11.0 1
B 23.0 1
C 9.0 1
;
run;
ods html;
PROC GLMPOWER data = d11_20;
class method;
model score = method;
weight CellWgt;
contrast "A vs. B" method 1 -1 0;
contrast "A vs. C" method 1 0 -1;
contrast "B vs. C" method 0 1 -1;
POWER stddev = 3.0 alpha = 0.05 power = 0.90 ntotal = .;
run; quit;
ods html close;
```

**【程序说明】**data 步定义影响因素各水平组指标的总均值的估计值及各水平组的样本含量分配比例;GLMPOWER 过程中的选项“stddev = 3.0”中的 3.0 为各水平组指标的总标准差的估计值,本例取三个水平组总标准差估计值的最大值带入计算,也可以同时考虑多个总标准差的估计值,将选项写成“stddev = 2.0 2.5 3.0”或“stddev = 2.0 to 3.0 by 0.5”的形式;程序中的数字需要根据实际情况进行修改。

具体地说,在数据步中:method 实际上指“试验因素”,若用 factor 更直观,它是一个字符型变量,其具体取值或水平可用一个字母或一个单词、甚至可用一句话(当字符个数超过 8 时,不能仅用一个美元符号,如最长为 20 个字符,可写成:method \$20.) 来表示“score”指各组的算术平均值,若用“mean”更直观,“CellWgt”指各组的样本含量比例或权重,本例取 1:1:1 根据实际需要,也可取其它比例关系,如 1:2:3;

在过程步中,有三个 Contrast 语句,相当于单因

素三水平之间的两两比较,分别表示 A 与 B 比较、A 与 C 比较、B 与 C 比较; POWER 语句中写了 4 个内容,可以写在同一行上,写成 4 行是为了看起来更清楚一些,前 3 行代表提供的 3 个已知条件,它们分别

为平均的或最大的样本标准差、 $\alpha$  值和  $1 - \beta$  的值;而“ntotal = .”表示需要估计总样本含量,即三组样本含量之总和。

【输出结果】

Computed N Total						
Index	Type	Source	Test DF	Error DF	Actual Power	N Total
1	Effect	method	2	6	0.989	9
2	Contrast	A vs. B	1	6	0.981	9
3	Contrast	A vs. C	1	141	0.900	144
4	Contrast	B vs. C	1	6	0.997	9

【结果解读】由计算结果可知:若要求通过单因素三水平设计一元定量资料的方差分析及两两比较有 90% 的把握发现三个组 SSS 评分的总体均值不完全相等,则每组只需 3 例患者,三组共需 9 例患者;但若要求通过单因素三水平设计一元定量资料的方差分析及两两比较有 90% 的把握发现三个组中任何两组的 SSS 评分的总体均值不相等,则每组需要 48 例患者,三组共需 144 例患者。

为什么会出现这样大的差异呢?问题的症结在于 A 组与 C 组的均值过于接近,它们之间的数值之差量为  $11 - 9 = 2$ ,要想把总体上存在如此小的“差距”都能发现出来,必须具有非常大的样本含量。

试想:若 A 组的均值不是 11,而是 16,可以计算得出:每组只需要 5 例,三组共需要 15 例即可。程序修改之后的输出结果如下:

Computed N Total						
Index	Type	Source	Test DF	Error DF	Actual Power	N Total
1	Effect	method	2	6	0.976	9
2	Contrast	A vs. B	1	12	0.922	15
3	Contrast	A vs. C	1	12	0.922	15
4	Contrast	B vs. C	1	6	0.997	9

【例 4】欲研究某降压药对高血压患者的疗效,拟对两个不同药物剂量组分别在用药前、用药后 1 周、2 周、3 周、4 周测量患者血压,假定两组间可识别的血压最小差异为 10 mmHg,欲推断不同剂量组间疗效的差异是否具有统计学意义,则每组需要多少患者( $\alpha = 0.05, 1 - \beta = 0.90$ )? 预试验已获知信息有: $\sigma_e^2 = 190.641$  条件相关系数  $\rho_c = 0.75, \sigma_u^2 = 457.333$ 。

```
( u005 + u02 ) * ( u005 + u02 ) / ( delta * delta ) ) + 1;
PUT #2 @15 '两组各需' n '例。';
RUN;
% MEND repeatmeasure;
DATA ;
% canshu;
% repeatmeasure( a b c d );
RUN;
```

【分析与解答】这是做具有一个重复测量因素的两因素设计一元定量资料均值比较之前进行样本含量估计的问题。所需的 SAS 程序如下。

【程序说明】第 2 行中“alpha = 0.05”中的 0.05 为可能犯第 I 类错误的概率,“beta = 0.10”中的 0.10 为可能犯第 II 类错误的概率,“k = 5”中的 5 为重复测量次数,“pc = 0.75”中的 0.75 为条件相关系数,“delta = 10”中的 10 为两组间可以识别的最小差异,“deltae = 190.641”中的 190.641 为重复测量误差,“deltau = 457.333”中的 457.333 为个体间差异的方差。

```
% MACRO canshu;
  alpha = 0.05; beta = 0.10; k = 5; pc = 0.75; delta = 10;
  deltae = 190.641; deltau = 457.333;
% MEND canshu;
% MACRO repeatmeasure( a b c d ); FILE PRINT;
u005 = put( ( -probit( alpha/2 ) ) / delta );
u02 = put( ( -probit( beta ) ) / delta );
n = int( ( 2 * ( deltau + ( 1 + ( k - 1 ) * pc ) * deltae / k ) *
  【输出结果】每组需要 129 例。
```

【输出结果】每组需要 129 例。

3 生存分析时样本含量估计

【例 5】某临床试验拟比较 TIPS 手术与脊髓改

道手术延长出血性食管曲张患者的生存时间有无差别, 试验组为 TIPS 手术组, 对照组为脊髓改道手术组, 随访期至少 1 年。根据既往研究可知对照组总体 1 年生存率是 45%, 预期试验组总体 1 年生存率将达到 65%, 采用 Log - Rank 检验比较两组生存率, 试估计样本含量 (采用双侧检验, 检验水准  $\alpha = 0.05$ , 检验效能  $1 - \beta = 0.85$ )。

【分析与解答】对照组生存率为 45%、试验组生存率为 65%, 对照组和试验组的终点事件发生率分别为:  $\pi_1 = 1 - 0.65 = 0.35$ ,  $\pi_c = 1 - 0.45 = 0.55$ 。这是对生存资料分析之前的样本含量估计问题。所需的 SAS 程序如下。

```
% let pt=0.35; % let pc=0.55; % let alpha=0.05; % let beta
=0.15;

data a6_1;
theta = log10(1 - &pt) / log10(1 - &pc);
theta1 = round(theta * 0.001);
e = ((theta + 1) / (theta - 1)) ** 2 * (probit(1 - &alpha/2)
+ probit(1 - &beta)) ** 2;
e1 = int(e) + 1;
n = 2 / (&pt + &pc) * e;
n1 = int(n) + 1;
file print;

PUT #3 @15 '风险比为 ' theta1 '; 两组终点事件发生总例
数不小于 ' e1 '; 两组所需总的样本含量为 ' n1 '。';

run;
```

【程序说明】程序当中 pt、pc、alpha、beta 分别代表试验组终点事件发生率  $\pi_1$ 、对照组终点事件发生率  $\pi_c$ 、检验水准  $\alpha$ 、犯假阴性错误的概率  $\beta$ 。

【输出结果】风险比为 0.539; 两组终点事件发生总例数不小于 101; 两组所需总的样本含量为 223。

说明: 样本含量与检验效能估计的内容非常丰富, 本文的目的仅仅希望起到抛砖引玉的作用。这方面的统计软件和书籍<sup>[6-7]</sup>很多, 读者需要时, 还可参阅有关文献。

## 参考文献

- [1] 张效嘉, 胡良平. 精神卫生科研如何严格遵守试验设计四原则之重复原则[J]. 四川精神卫生, 2016, 29(4): 303 - 306.
- [2] SAS Institute Inc. SAS/STAT 9.3 User's Guide[M]. Cary, NC: SAS Institute Inc, 2011: 3361 - 3400, 5729 - 5962.
- [3] 胡良平. 统计学三型理论在实验设计中的应用[M]. 北京: 人民军医出版社, 2006: 215 - 265.
- [4] 胡良平. 科研设计与统计分析[M]. 北京: 军事医学科学出版社, 2012: 136 - 227.
- [5] 胡良平. SAS 常用统计分析教程[M]. 北京: 电子工业出版社, 2015: 235 - 252.
- [6] Shein - Chung Chow, Jun Shao, Hansheng Wang. Sample Size Calculations in Clinical Research[M]. 2nd Edition. Chapman & Hall/CRC, 2008: 49 - 408.
- [7] Thomas P. Ryan. Sample Size Determination and Power[M]. Wiley, 2013: 17 - 352.

(收稿日期: 2016 - 10 - 11)

(本文编辑: 吴俊林)