Methodology 方法学

Estimation of sample size and testing power (Part 6)

Liang-ping Hu, Xiao-lei Bao, Xue Guan, Shi-guo Zhou
Consulting Center of Biomedical Statistics, Academy of Military Medical Sciences, Beijing 100850, China

ABSTRACT: The design of one factor with k levels ($k \geqslant 3$) refers to the research that only involves one experimental factor with k levels ($k \geqslant 3$), and there is no arrangement for other important non-experimental factors. This paper introduces the estimation of sample size and testing power for quantitative data and qualitative data having a binary response variable with the design of one factor with k levels ($k \geqslant 3$).

KEYWORDS: statistics, medical; research design; sample size; testing power; parametric estimation

The design of one factor with k levels ($k \geqslant 3$) refers to the research that only involves one experimental factor with k levels ($k \geqslant 3$) and there is no arrangement for other important non-experimental factors because the researchers want to balance the influence of the important non-experimental factors by randomized grouping[1]. If the experimental factor is independent among research subjects, then the subjects can be divided into k groups by complete randomization; otherwise, the research subjects should be selected from k subpopulations. In this paper, we will introduce the estimation of sample size and testing power for quantitative data and qualitative data having a binary response variable with the design of one factor with k levels ($k \geqslant 3$).

1 Quantitative data with the design of one factor with k levels ($k \geqslant 3$)

1.1 Estimation of sample size

1.1.1 Formula The formula for sample size estimation for quantitative data with the design of one factor with k levels ($k \geqslant 3$) is as follows[1-3]:

$$n = \psi^2 \left(\sum_{i=1}^{k} \hat{\sigma}_i^2 / k \right) \Big/ \left[\sum_{i=1}^{k} (\hat{\mu}_i - \bar{\hat{\mu}})^2 / (k-1) \right] \quad (1)$$

In formula (1), n stands for the sample size of each group, which is required to be equal; $\hat{\mu}$ and $\hat{\sigma}_i$ refer to the estimated population mean and the estimated population standard deviation of the group i; $\bar{\hat{\mu}} = \sum_{i=1}^{k} \hat{\mu}_i / k$, in which k stands for the

Open Access 开放获取

DOI: 10.3736/jcim20120308
http://www.jcimjournal.com

Hu LP, Bao XL, Guan X, Zhou SG. Estimation of sample size and testing power (Part 6). *J Chin Integr Med*. 2012; 10(3): 298-302.
胡良平, 鲍晓蕾, 关雪, 周诗国. 样本量估计与检验效能分析(六). 中西医结合学报. 2012; 10(3): 298-302.

Received December 29, 2011; accepted December 31, 2011; published online March 15, 2012.
Full-text LinkOut at PubMed. Journal title in PubMed: *Zhong Xi Yi Jie He Xue Bao*.

基金项目: 国家科技重大专项"重大新药创制"资助项目 (No. 2009ZX09502-028)
Correspondence: Prof. Liang-ping Hu; Tel: 010-66931130; E-mail: lphu812@sina.com

ISSN 1672-1977. Published by JCIM Press, Shanghai, China.

number of the groups; $\psi=\sqrt{\frac{\delta}{v_1}}$, in which δ is the non-central parameter of the non-central F distribution under certain conditions; and $v_1=k-1$ is the degree of freedom of the experimental factor with the design of one factor with k levels ($k\geqslant 3$). The value of ψ can be obtained by SAS function, or by the ψ table in reference books. The iterative algorithm is adopted for estimation. First, compute the ψ value by using the known information which includes α, β, $v_1=k-1$ and $v_2=\infty$ or by checking the ψ table, and apply it to formula (1) in order to get $n_{(1)}$; then compute the ψ value by using the known information including $v_1=k-1$ and $v_2=k(n_{(1)}-1)$ or by checking the ψ table and apply it to formula (1) in order to get $n_{(2)}$, and so on... The computation stops when the nearest two results are stable, and this is the required sample size.

1.1.2 Example 1 Three methods were adopted in treatment of patients suffering from poststroke depression. The neurological rehabilitation status was observed. The means of the Scandinavian Stroke Scale (SSS) score after treatment by the three methods were estimated to be 11.0, 10.0 and 9.0, respectively, and the standard deviations were 3.0, 3.0 and 2.0, respectively. If the difference of the three methods was required to be statistically significant, how many patients were needed for each group (α and β were set to be 0.05 and 0.10, respectively)?

Analysis: Example 1 deals with the sample size estimation of the population mean comparison for quantitative data with the design of one factor with three levels. Based on the given information, we can calculate that $\hat{\bar{\mu}}=(11.0+10.0+9.0)/3=10.0$, $\sum\hat{\sigma}_i^2=3.0^2+3.0^2+2.0^2=22.0$, and $\sum(\hat{\mu}-\hat{\bar{\mu}})^2=(11.0-10.0)^2+(10.0-10.0)^2+(9.0-10.0)^2=2.0$. The given information also includes $v_1=3-1=2$, $v_2=\infty$, $\alpha=0.05$ and $\beta=0.10$, thus we know that $\psi_{0.05,0.10,2,\infty}=2.52$ by checking the ψ table, and $n_{(1)}=(2.52)^2\times(22.0/3)/[2.0/(3-1)]\approx 46.6$. Let $n_{(1)}=47$, based on the known information $v_1=3-1=2$, $v_2=3\times(47-1)=138$, $\alpha=0.05$ and $\beta=0.10$, we get $\psi_{0.05,0.10,2,138}\approx\psi_{0.05,0.10,2,120}=2.55$ and $n_{(2)}=(2.55)^2\times(22.0/3)/[2/(3-1)]\approx 47.7$. Let $n_{(2)}=48$, based on the known information $v_1=3-1=2$, $v_2=3\times(48-1)=141$, $\alpha=0.05$ and $\beta=0.10$, we get $\psi_{0.05,0.10,2,141}\approx\psi_{0.05,0.10,2,120}=2.55$ and $n_{(3)}=(2.55)^2\times(22.0/3)/[2/(3-1)]\approx 47.7$. Let $n_{(3)}=48$, and $n=48$ is the finial sample size; that is, each group needs 48 patients, and the total sample size would be 144.

Besides, the following two SAS programs can be applied to estimate the sample size. Bellow is the first SAS program:

```
data example_1;
input method $ score CellWgt;
datalines;
A 11.0 1
B 10.0 1
C  9.0 1
;
run;
ods html;
PROC GLMPOWER data=example_1;
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 = 2.7
alpha  = 0.05
ntotal = .
power  = 0.90;
run;
ods html close;
quit;
```

Program explanation: In the beginning of the program, the estimated population means and the sample size ratio of the three groups are specified. The option "stddev = 2.7" in the glmpower procedure refers to the mean of the estimated population standard deviations of the three groups. We

can also input the different estimated population standard deviations by changing the option into "stddev = 2.0 2.5 3.0" or "stddev = 2.0 to 3.0 by 0.5". The option "alpha = 0.05" specifies the value of α; "ntotal = ." requires to estimate the population sample size; "power = 0.90" specifies the testing power $1-\beta$ to be 0.90. The parameter values in the program can be altered in similar situations.

Output and explanation:

Fixed scenario elements

Dependent variable	Score
Weight variable	CellWgt
Alpha	0.05
Error standard deviation	2.7
Nominal power	0.9

Computed *N* Total

Index	Type	Source	Test degree of freedom	Error degree of freedom	Actual power	*N* total
1	Effect	method	2	141	0.906	144
2	Contrast	A *vs.* B	1	459	0.900	462
3	Contrast	A *vs.* C	1	114	0.900	117
4	Contrast	B *vs.* C	1	459	0.900	462

The above results show that if the testing power to find out that the population means of the SSS score of the three groups are unequal and required to be 90%, 48 patients are needed for each group, namely, a total of 144 patients are needed. Furthermore, if the testing power to find out that the population means of the SSS score of each two groups are unequal and required to be 90%, 154 patients are needed for each group, that is, a total of 462 patients are needed.

The other SAS program is as follows:

```
ods html;
PROC POWER;
onewayanova
groupmeans =11.0 | 10.0 | 9.0
stddev = 2.7
groupweights = (1 1 1)
alpha = 0.05
Ntotal =.
power =0.90;
run;
ods html close;
quit;
```

Program explanation: The option "groupmeans =11.0|10.0|9.0" specifies the estimated population means of the three groups; "stddev = 2.7" specifies the estimated population standard deviations of the three groups; "groupweights = (1 1 1)" specifies the sample size ratio of the three groups. The parameter values can be modified in similar situations.

Output and explanation:

Overall *F* test for one-way ANOVA

Fixed Scenario Elements	
Method	Exact
Alpha	0.05
Group means	11 10 9
Standard deviation	2.7
Group weights	1 1 1
Nominal power	0.9

Computed *N* Total

Actual power	*N* total
0.906	144

The results show that if the testing power to find out that the population means of the SSS score of the three groups are unequal and required to be 90%, 48 patients are needed for each group, that is, a total of 144 patients are needed.

1.2 Estimation of testing power The following example is used to demonstrate the estimation of testing power for quantitative data with the design of one factor with k levels ($k \geqslant 3$).

Example 2 A researcher expected to examine the difference of the serum soluble CD8 antigen level (U/mL) between patients suffering from mild and severe aplastic anemia and healthy people in order to reflect the extent of hematopoietic dysfunction caused by the disorder of the immune status. The researcher selected 10 subjects from the three types of population respectively (patients suffering from mild aplastic anemia, patients suffering from severe aplastic anemia and healthy people), and examined the CD8 antigen level. Below was the result ($\overline{x} \pm s$): the healthy group 290±174, the mild group 658±155 and the severe group 763±127. The researcher analyzed the data by adopting the analysis of variance of quantitative data with the design of one factor with three levels. Estimate the testing power.

Analysis: Example 2 involves the testing power estimation for quantitative data with the design of one factor with three levels. The following SAS program can be applied.

```
proc power;
onewayanova
groupmeans =290| 658 | 763
stddev = 174 155 127
groupweights = (1 1 1)
alpha = 0.05
ntotal =30
power =.;
run;quit;
```

Program explanation: The option "groupmeans =290| 658 | 763" specifies the estimated population mean of each group; "stddev = 174 155 127" specifies the estimated population standard deviation of each group; "groupweights = (1 1 1)" specifies the sample size ratio of each group.

Output and explanation:

Overall *F* test for one-way ANOVA

Fixed scenario elements	
Method	Exact
Alpha	0.05
Group means	290 658 763
Group weights	1 1 1
Total sample size	30

Computed power

Index	Standard deviation	Power
1	174	>0.999
2	155	>0.999
3	127	>0.999

The result shows that the testing power was > 0.999.

2 Qualitative data with the design of one factor with *k* levels (*k*≥3) having a binary response variable

2.1 Sample size estimation of multirate comparison

2.1.1 Formula

$$n=\frac{\lambda}{2(\arcsin\sqrt{p_{max}}-\arcsin\sqrt{p_{min}})^2} \quad (2)$$

In formula (2), n refers to the sample size of each group, which is required to be equal; p_{max} and p_{min} stand for the maximum rate and the minimum rate. When the difference of p_{max} and p_{min} (p_d) is known, $p_{max}=0.5+p_d/2$, $p_{min}=0.5-p_d/2$; λ is the non-central parameter δ_{χ^2} of the non-central χ^2 distribution under certain conditions (df=$k-1$), which can be computed by SAS function or by checking the λ table in reference books. k stands for the number of the groups.

2.1.2 Example 3 Three drugs used to expel the intestinal worm parasite were applied to a pilot test. The stool examination showed that the egg negative conversion rates of drug A, drug B and drug C were 80%, 85% and 95%, respectively. How many patients were needed for each group in the clinical trial?

Analysis: Example 3 deals with the sample size estimation of multirate comparison. The needed SAS program is as follows:

Program explanation: In the first line, "alpha =0.05" specifies the probability of making type Ⅰ error; "beta=0.10" specifies the probability of making type Ⅱ error; "$k=3$" specifies the number of the groups. In the second line, "$p=$ 0.80,0.85,0.95" specifies the egg negative conversion rate of each group. The value of λ can be obtained by checking the λ table or by invoking the SAS function. The values of the parameters can be altered in similar situations.

```
%let alpha=0.05; %let beta=0.10; %let k=3; %let p=
0.80,0.85,0.95;
data example_3;
p1=max(&p);
p2=min(&p);
lamda=CNONCT(CINV(1-&alpha,&k-1),&k-1,&beta);
n=ceil(lamda/(2 * (arsin(sqrt(p1))-arsin(sqrt(p2))) * *
2));
file print;
PUT #3 @10 ''n 'patients were needed for each group.';
run;
```

The result: 112 patients are needed for each group.

2.2 Testing power estimation of multirate comparison

2.2.1 Formula First, compute the value of λ by formula (2), then invoke the SAS function to solve the value of β in function λ = CNONCT (CINV($1-\alpha$, $k-1$), $k-1$, β) and at last compute the testing power ($power=1-\beta$).

2.2.2 Example 4 In example 3, suppose there are overall 300 patients in the clinical trial and each group has 100 patients. Estimate the testing power.

Analysis: Example 4 deals with the testing power estimation of multirate comparison. The needed SAS program is as follows:

```
%let alpha=0.05; %let k=3; %let n=100; %let p=0.80,
0.85,0.95;
data example_4;
p1=max(&p);
p2=min(&p);
lamda=&n * 2 * (arsin(sqrt(p1))-arsin(sqrt(p2))) * * 2;
do beta=0.001 to 0.999 by 0.001;
lamda_1=CNONCT(CINV(1-&alpha,&k-1),&k-1,beta);
output;
if abs(lamda_1-lamda)/lamda<=0.01 then goto ok;
end;
ok:power=1-beta;
file print;
PUT #3 @10 'The testing power reaches 'power'.';
run;
```

Program explanation: The value of beta in the statement "do…, …end" can not be 0 or 1, otherwise, it will be unable to calculate the value of lamda_1, which may lead to a wrong result. The statement "if abs(lamda_1-lamda)/lamda<=0.01 then goto ok;" requires to set a reasonable error (for instance, the maximum error is set to be 1%), otherwise the result may also turn out to be wrong.

Output: The testing power reaches 0.867.

REFERENCES

1 Hu LP. The application of statistical triple-type theory

in experimental designs. Beijing: People's Military Medical press. 2006: 57, 222-224, 318-319. Chinese.
胡良平. 统计学三型理论在实验设计中的应用. 北京: 人民军医出版社. 2006: 57, 222-224, 318-319.
2 Hu LP. SAS research design and statistical analysis. Beijing: People's Medical Publishing House. 2010: 200-202, 208-209. Chinese.
胡良平. SAS实验设计与统计分析. 北京: 人民卫生出版社. 2010: 200-202, 208-209.
3 Hu LP, Gao H. Guidance of statistical analysis by SAS. Beijing: Publishing House of Electronics Industry. 2010: 700-701. Chinese.
胡良平, 高辉. SAS统计分析教程. 北京: 电子工业出版社. 2010: 700-701.

样本量估计与检验效能分析(六)

胡良平, 鲍晓蕾, 关雪, 周诗国
军事医学科学院生物医学统计学咨询中心, 北京 100850

摘要: 所谓单因素多水平设计,是指试验中仅涉及一个具有 k 个水平($k \geqslant 3$)的试验因素,未对其他任何重要非试验因素进行有计划的安排。本文向读者介绍单因素多水平设计一元定量资料与结果变量为二值变量的单因素多水平设计一元定性资料的样本含量与检验效能估计。

关键词: 统计学, 医学; 研究设计; 样本大小; 检验效能; 参数估计