Methodology 方法学

Estimation of sample size and testing power (Part 7)

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

ABSTRACT: Two-factor factorial design refers to the research involving two experimental factors and the number of the experimental groups equals to the product of the levels of the two experimental factors. In other words, it is the complete combination of the levels of the two experimental factors. The research subjects are randomly divided into the experimental groups. The two experimental factors are performed on the subjects at the same time, meaning that there is no order. The two experimental factors are equal during statistical analysis, that is to say, there is no primary or secondary distinction, nor nested relation. This article introduces estimation of sample size and testing power of quantitative data with two-factor factorial design.

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

Two-factor factorial design is frequently adopted when the research involves two experimental factors, which are performed on the research subjects at the same time and are equal, and have level 1 interaction which is required to be examined. When there is only one quantitative index, it is called univariate quantitative data with two-factor factorial design, which can be analyzed by the analysis of variance of univariate quantitative data with two-factor factorial design[1]. This article introduces the estimation of sample size and testing power of univariate quantitative data with two-factor factorial design. Since the estimation of sample size and testing power of two-factor factorial design is complicated, this article will introduce it through examples when performing the analysis of variance by SAS.

1 Estimation of sample size

1.1 Example 1 A researcher planed to conduct an animal experiment to study the influence of copper (Cu) and vitamin E (VE) on liver injury caused by carbon tetrachloride (CCl_4). Thirty healthy male Wistar rats were chosen for pilot test and were equally divided into 10 groups. How many rats were needed to perform the analysis of variance of two-factor factorial design in the formal experiment if the testing power was required to be 80%[2]?

1.2 Analysis Step 1: Based on the information of the pilot test, estimate the population means and the population standard deviations of superoxide

Open Access 开放获取

DOI: 10.3736/jcim20120405
http://www.jcimjournal.com

Hu LP, Bao XL, Guan X. Estimation of sample size and testing power (Part 7). *J Chin Integr Med*. 2012; 10(4): 380-383.
胡良平, 鲍晓蕾, 关雪. 样本量估计与检验效能分析(七). 中西医结合学报. 2012; 10(4): 380-383.

Received February 10, 2012; accepted February 22, 2012; published online April 15, 2011. 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.

dismutase (SOD) under all the experimental conditions. The corresponding SAS program named *exam*1_1.*sas* is as follows:

```
data step1;
do Cu=0.00,0.05,0.10,0.20,0.40;
do VE=0,150;
do i=1 to 3;
input SOD@@; output;
end;end;end;
cards;
343.8  331.6  318.1  465.2  457.7  449.5
430.0  417.0  419.3  584.5  570.0  576.7
448.1  413.1  454.4  604.3  531.4  605.0
443.7  412.3  468.7  485.6  516.4  485.9
474.6  455.5  464.4  509.8  552.0  498.6
;
run;
ods html;
proc means mean std;
var SOD;
class Cu VE;
run;
ods html close;
```

Program explanation: First, create a new data set named *step*1; then invoke the MEANS procedure to estimate the means and the standard deviations.

Below is the main output:

Response Variable: SOD

Cu	VE	Sample size	Mean	Standard deviation
0	0	3	331.166 666 7	12.855 478 7
	150	3	457.466 666 7	7.852 600 4
0.05	0	3	422.100 000 0	6.937 578 8
	150	3	577.066 666 7	7.256 950 7
0.1	0	3	438.533 333 3	22.250 018 7
	150	3	580.233 333 3	42.292 355 5
0.2	0	3	441.566 666 7	28.260 455 3
	150	3	495.966 666 7	17.696 421 5
0.4	0	3	464.833 333 3	9.557 370 6
	150	3	520.133 333 3	28.159 782 2

Step 2: Estimate the sample size based on the result of step 1. The needed SAS program named *exam*1_2.*sas* is as follows:

```
data step2;
do Cu=0.00,0.05,0.10,0.20,0.40;
do VE=0,150;
input mean @@; output;
end;end;
cards;
331.1666667 457.4666667
422.1000000 577.0666667
438.5333333 580.2333333
441.5666667 495.9666667
464.8333333 520.1333333
;
run;
ods html;
proc glmpower data=step2;
class Cu VE;
model mean=Cu|VE;
power
stddev=6 to 42   by 12
ntotal=.
power=0.80;
run;quit;
ods html close;
```

Program explanation: First, specify the population means under all the experimental conditions. The "glmpower" procedure performs prospective power analysis. The statement "model mean=Cu | VE" can also be written as "model mean= Cu VE Cu * VE". The option "stddev=6 to 42 by 12" specifies 4 population standard deviations, namely, 6, 18, 30 and 42 since step 1 has already computed the minimum standard deviation as 6.9, and the maximum as 42.3. The option "power=0.80" designates 0.8 as the testing power. The above parameter values can be altered under specific situations.

The main output is as follows:

Fixed Scenario Elements

Dependent variable	Mean
Nominal Power	0.8
Alpha	0.05

Related Articles 推荐阅读

卞兆祥, Moher D, 李幼平, 吴泰相, Dagenais S, 郑颂华, 李静, 李廷谦. 清晰报告中医药随机对照试验中的干预措施. 中西医结合学报. 2008; 6(7): 661-667.

Bian ZX, Moher D, Li YP, Wu TX, Dagenais S, Cheng CW, Li J, Li TQ. Precise reporting of traditional Chinese medicine interventions in randomized controlled trials. *J Chin Integr Med*. 2008; 6(7): 661-667.

Free full text available at http://www.jcimjournal.com/FullText2.aspx?articleID=167219772008070661

卞兆祥, 李幼平, Moher D, Dagenais S, 刘良, 吴泰相, 缪江霞, 关家伦, 宋丽. 提高中草药随机对照试验的质量Ⅰ: 临床试验设计和方法学. 中西医结合学报. 2006; 4(2): 120-129.

Bian ZX, Li YP, Moher D, Dagenais S, Liu L, Wu TX, Miao JX, Kwan AKL, Song L. Improving the quality of randomized controlled trials in Chinese herbal medicine, Part Ⅰ: clinical trial design and methodology. *J Chin Integr Med*. 2006; 4(2): 120-129.

Free full text available at http://www.jcimjournal.com/FullText2.aspx?articleID=502

More free related articles at http://www.jcimjournal.com/FullText2.aspx?articleID=jcim20120405

Computed N Total

Index	Source	Std Dev	Test DF	Error DF	Actual Power	N Total
1	Cu	6	4	10	>0.999	20
2	Cu	18	4	10	>0.999	20
3	Cu	30	4	10	0.984	20
4	Cu	42	4	10	0.812	20
5	VE	6	1	10	>0.999	20
6	VE	18	1	10	>0.999	20
7	VE	30	1	10	>0.999	20
8	VE	42	1	10	>0.999	20
9	Cu* VE	6	4	10	>0.999	20
10	Cu* VE	18	4	10	0.940	20
11	Cu* VE	30	4	20	0.812	30
12	Cu* VE	42	4	50	0.877	60

The above result presents the needed sample sizes with different standard deviations. When the population standard deviation is 42, the testing power to infer that Cu has a protective effect on liver injury caused by CCl_4 is 81.2%, and the testing power to infer that VE has a protective effect on liver injury caused by CCl_4 is 99.9% as long as the sample size reaches 20. When the sample size is 60, the testing power to infer that there is interaction between Cu and VE is 87.7%. Therefore, when the population standard deviation is 42, 60 rats are needed; that is to say, each group needs 6 rats at least since there are 10 groups.

2 Estimation of testing power

2.1 Example 2 In example 1, suppose that the researcher increased the sample size to 60 and conducted the additional experiments. The researcher wondered whether the testing power was sufficient if adopting the analysis of variance of univariate quantitative data with two-factor factorial design[2].

2.2 Analysis Step 1, based on the information of the pilot test, estimate the population means and the population standard deviations of SOD under all the experimental conditions. See program *exam*1_1*.sas* for reference. The main output is as follows:

Response Variable: SOD

Cu	VE	Sample size	Mean	Standard deviation
0	0	6	341.450 000 0	15.695 445 2
	150	6	463.766 666 7	9.612 006 4
0.05	0	6	413.183 333 3	11.141 708 4
	150	6	573.183 333 3	19.615 750 5
0.1	0	6	423.866 666 7	28.221 386 7
	150	6	595.783 333 3	45.698 639 7
0.2	0	6	453.133 333 3	61.289 237 8
	150	6	495.883 333 3	16.018 915 9
0.4	0	6	459.716 666 7	25.860 581 3
	150	6	521.166 666 7	36.075 291 6

Step 2: Estimate the testing power based on the above results. The corresponding SAS program named *exam*1_2*.sas i*s as follows:

```
data step3;
do Cu=0.00,0.05,0.10,0.20,0.40;
do VE=0,150;
input mean @@; output;
end;end;
cards;
341.4500000 463.7666667
413.1833333 573.1833333
423.8666667 595.7833333
453.1333333 495.8833333
459.7166667 521.1666667
;
run;
ods html;
proc glmpower data=step3;
class Cu VE;
model mean=Cu|VE;
power
stddev=9.6 45.7
ntotal=60
power=.;
run;quit;
ods html close;
```

Program explanation: First, specify the population means under all the experimental conditions. The glmpower procedure performs prospective power analysis. The option "stddev=9.6 45.7" designates 9.6 and 45.7 as the minimum and the maximum population standard deviations. The option "ntotal=60" designates 60 as the total sample size.

Below is the main output:

Fixed Scenario Elements

Dependent variable	Mean
Total sample size	60
Alpha	0.05
Error degrees of freedom	50

Computed Power

Index	Source	Std Dev	Test DF	Power
1	Cu	9.6	4	>0.999
2	Cu	45.7	4	>0.999
3	VE	9.6	1	>0.999
4	VE	45.7	1	>0.999
5	Cu* VE	9.6	4	>0.999
6	Cu* VE	45.7	4	0.936

The above result shows that when the sample size is 60, the testing power to infer that Cu has a protective effect on liver injury caused by CCl_4 reaches 99.9% and 99.9% (S=9.6 and S=45.7), and the testing power to infer that VE has a protective effect on liver injury caused by CCl_4 reaches 99.9% and 99.9% (S=9.6 and S=45.7). When the population standard deviation is 9.6, the testing power to infer that there is interaction between Cu and VE reaches 99.9%. When the population standard deviation is 45.7, the testing power reaches 93.65%.

REFERENCES

1 Hu LP. The application of statistical triple-type theory in experimental designs. Beijing: People's Military Medical Press. 2006: 94. 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: 94.

2 Hu LP. SAS research design and statistical analysis. Beijing: People's Medical Publishing House. 2010: 202-204, 223-225. Chinese.
胡良平. SAS 实验设计与统计分析. 北京: 人民卫生出版社. 2010: 202-204, 223-225.

样本量估计与检验效能分析(七)

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

摘要：所谓两因素析因设计，是指试验中涉及两个试验因素，不同的试验条件数等于两个试验因素的水平数的乘积，即由试验因素的水平全面组合而成。全部受试对象被完全随机地分配到各试验条件组中去。做试验时，全部试验因素同时施加(即无先后顺序之分)；对资料进行统计分析时，试验因素地位平等(即不存在主次之分，也不存在嵌套关系)。本文向读者介绍采用两因素析因设计一元定量资料方差分析处理定量资料时，在试验之前如何估计样本含量与检验效能。

关键词：统计学，医学；研究设计；样本大小；检验效能；参数估计