

Methods and analysis of realizing randomized grouping

Liang-ping Hu, Xiao-lei Bao, Qi Wang

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

Abstract: Randomization is one of the four basic principles of research design. The meaning of randomization includes two aspects: one is to randomly select samples from the population, which is known as random sampling; the other is to randomly group all the samples, which is called randomized grouping. Randomized grouping can be subdivided into three categories: completely, stratified and dynamically randomized grouping. This article mainly introduces the steps of complete randomization, the definition of dynamic randomization and the realization of random sampling and grouping by SAS software.

Keywords: statistics, medical; research design; random allocation; SAS program

In scientific studies, researchers often confront various problems involving randomization, for instance, how to randomly select research subjects for sampling survey, and how to randomly divide subjects into groups properly in experimental design and clinical design, etc. There are many solutions to realize randomization. This article will focus on how to realize randomized grouping and sampling.

1 General methods to realize randomization

There are many methods to realize randomization, including drawing lots, looking up the random number table or the random arrangement table, looking up the pseudo-random number table produced by the computer or directly invoking computer programs^[1].

2 General steps to realize complete randomization

Step 1: Number all the subjects, and write down the numbers orderly.

Step 2: Set the rules for grouping beforehand. For example, if two groups are needed in design, divide the subjects with even numbers into the experimental group and the subjects with odd numbers into the control group (or the opposite). If three groups are needed in design, put the subjects with numbers divisible by 3 into the first group, and the subjects not divisible by 3 but with remainder 1 into the second group and the remaining subjects into the third group. Of course, other rules for grouping are also acceptable as long as the rules are stipulated beforehand and strictly followed.

 Open Access 开放获取	 Related Articles 推荐阅读
<p>DOI: 10.3736/jcim20110703 http://www.jcimjournal.com</p> <p>Hu LP, Bao XL, Wang Q. Methods and analysis of realizing randomized grouping. <i>J Chin Integr Med.</i> 2011; 9(7): 711-714.</p> <p>胡良平, 鲍晓蕾, 王琪. 随机化分组的方法. <i>中西医结合学报.</i> 2011; 9(7): 711-714.</p> <p>Received May 11, 2011; accepted May 16, 2011; published online July 15, 2011.</p> <p>Full-text LinkOut at PubMed. Journal title in PubMed: <i>Zhong Xi Yi Jie He Xue Bao.</i></p> <p>Correspondence: Prof. Liang-ping Hu; Tel: 010-66931130; E-mail: lphu812@sina.com</p>	<p>刘兆兰, 刘宝全, 李兰花, 刘建平. 动态随机分组方法介绍及应用. <i>中西医结合学报.</i> 2011; 9(3): 246-251.</p> <p>Liu ZL, Liu BQ, Li LH, Liu JP. Introduction of dynamic balancing randomization method and its application. <i>J Chin Integr Med.</i> 2011; 9(3): 246-251.</p> <p>Full text available at http://www.jcimjournal.com/FullText2.aspx?articleID=jcim20110304</p> <p>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. <i>中西医结合学报.</i> 2006; 4(2): 120-129.</p> <p>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. <i>J Chin Integr Med.</i> 2006; 4(2): 120-129.</p> <p>Full text available at http://www.jcimjournal.com/FullText2.aspx?articleID=502</p> <p>More related articles at http://www.jcimjournal.com/FullText2.aspx?articleID=jcim20110703</p>

Step 3: Copy the random numbers orderly from the above three random number tables and put them below the subject numbers. Abandon those not meeting the requirement, for instance, the random number exceeds the maximum subject number.

Step 4: Perform the grouping according to the prior rules. When the sample sizes in each group are unequal, random adjustment is recommended to make them equal or close (from the view of statistical computation, the error is relatively smaller when the sample sizes of each group are equal). If the random arrangement table is adopted, the grouping result usually makes the sample size of each group equal. Of course, with the help of computer programs, randomized grouping will be easily achieved.

3 Introduction of a simple randomized grouping method

In clinical studies, patients are usually assigned to the experimental group and the control group. Some researchers tend to divide the patients who come first into the experimental group and those who come later into the control group, which is not scientific because the patients are probably not balanced on some important non-experimental factors such as disease state, disease duration and so on. It is possible that most of the patients who come to the hospital in a certain time suffer a severe disease while those who come to the hospital in another time suffer a relatively mild disease. Therefore, if the researcher performs the cross-over grouping, that is, divide the first-coming patient into the experimental group, then the second-coming patient into the control group, it is possible that the patients in the two groups differ considerably on some important non-experimental factors, which may cause the two groups to be unbalanced. Here we introduce a simple randomized grouping method, of which the grouping principle is based on the smallest unbalanced index^[2].

Firstly, select several important non-experimental factors based on specialty. Suppose one of the important non-experimental factors is gender (male and female) and the other is disease state (mild, medium and severe). Put the first-coming two patients into the experimental group and the control group respectively and record their genders and disease states. Score 1 for every appearance of each level of the two factors, compute the absolute differences of the two levels of each factor, and sum up the absolute differences to be the unbalanced index of the two patients on the two important non-experimental factors. When a third patient comes, add his score respectively to the experimental group as well as the control group, and assign him to the group with the smaller unbalanced index. Apply the same method to future patients till the sample size reaches the requirement.

For instance, assume that disease state (mild, medium and severe), disease duration (short and long) and daily exercise amount (small, large) are three important non-experimental factors that need to be considered for the grouping of patients suffering from periarteritis of the shoulder. Suppose both the experimental group and the control group have one patient with basic information as per that shown in Table 1. When a third patient comes, with severe disease state, short disease duration and a large amount of daily exercise, which group should he be divided into?

Table 1 Basic information of the first two patients and the balance situation after grouping

Important non-experimental factors	Score		
	Experimental group	Control group	Absolute difference
Disease state			
Mild	0	1	1
Medium	1	0	1
Severe	0	0	0
Disease duration			
Short	0	1	1
Long	1	0	1
Daily exercise amount			
Small	1	1	0
Large	0	0	0
Total	3	3	4*

* The total value 4 in table 1 is the unbalanced index.

The third patient should be grouped into the experimental group and the control group respectively. The result is shown in Tables 2 and 3.

Since the unbalanced index after grouping the third patient into the experimental group and the control group is 5 and 7, respectively, he or she should be grouped into the experimental group. This grouping method also applies to other new patients. If there are k ($k \geq 2$) groups, put the new patient respectively in all the experimental groups and the control group, compute the unbalanced indexes of the total $2 \times (k - 1)$ tables (each

Table 2 Balance situation if the third patient is grouped into the experimental group

Important non-experimental factors	Score		
	Experimental group	Control group	Absolute difference
Disease state			
Mild	0	1	1
Medium	1	0	1
Severe	1	0	1
Disease duration			
Short	1	1	0
Long	1	0	1
Daily exercise amount			
Small	1	1	0
Large	1	0	1
Total	6	3	5

Table 3 Balance situation if the third patient is grouped into the control group

Important non-experimental factors	Score		
	Experimental group	Control group	Absolute difference
Disease state			
Mild	0	1	1
Medium	1	0	1
Severe	0	1	1
Disease duration			
Short	0	2	2
Long	1	0	1
Daily exercise amount			
Small	1	1	0
Large	0	1	1
Total	3	6	7

new patient is supposed to be grouped into an experimental group and the control group to compute an unbalanced index) and the grouping plan corresponding to the smallest unbalanced index should be adopted.

4 Realization of randomized grouping by SAS

4.1 Example 1 A total of 100 mice are numbered

Mouse	Group	Mouse	Group	Mouse	Group	Mouse	Group	Mouse	Group
2	A	44	A	84	A	19	B	61	B
4	A	45	A	85	A	20	B	63	B
...
39	A	81	A	17	B	57	B	99	B
42	A	82	A	18	B	58	B	100	B

“Mouse” in the above table stands for the number of mice, and “Group” stands for the two groups.

5 Realization of random sampling by SAS

5.1 Example 2 The basic information of 20 patients numbered 1 to 20 is shown in Table 4. Select 10 patients as research subjects by simple randomization.

Table 4 Basic information of 20 patients

Patient number	Gender	Age (years)	Patient number	Gender	Age (years)
1	Female	60	11	Male	58
2	Female	64	12	Male	63
3	Male	37	13	Female	23
4	Female	57	14	Female	37
5	Female	41	15	Female	20
6	Female	31	16	Female	33
7	Male	60	17	Female	39
8	Male	64	18	Male	40
9	Male	58	19	Female	49
10	Male	16	20	Female	42

The program to realize simple random sampling is as follows:

from 1 to 100 and divided into 2 experimental groups (A and B) by complete randomization.

```
proc plan;
  factors i=100;
  output out=a;
run;
data b c;
  set a;
  mouse=_n_;
  if i<=50 then do;
    group='甲'; output b; drop i;
  end;
else do;
  group='乙'; output c; drop i;
end;
run;
```

```
data d;
  set b c;
run;
ods html;
proc print noobs;
run;
ods html close;
```

4.2 Explanation of the program Firstly, invoke the procedure PLAN to generate 100 random numbers and output the result to data set “a”. Then, divide the mice into two groups based on random numbers and put the result into data set “b” and “c”. Lastly, merge data set “b” and “c” into a new data set “d”, and output the result of randomized grouping by the procedure PRINT.

4.3 Output of the result The output of the result by SAS is shown as follows.

```
data a;
  input id sex $ age;
cards;
1 F 60
2 F 64
... ..
20 F 42
;
run;
```

```
ods html;
proc surveyselect data = a
  method=srs n=10 out=b;
run;
proc print data=b;
run;
ods html close;
```

5.2 Explanation of the program Firstly, create data set “a”. Then, invoke the procedure SURVEYSELECT to realize random sampling. The option “data=” specifies the input data set for sampling; “method=” specifies the method for random sampling; “srs” stands for simple random sampling; “n=” specifies the size of the expected sample, which here can be replaced by “rate=” to assign the sample rate; “out=” specifies the output data set including the selected subjects (the option “rep = number” can be added to speculate the replication times of sampling). Lastly, the procedure PRINT is used to output data set “b”. “ods html” and “ods html close” request SAS to output the result by web page.

The output of the result is as follows:

Selection method	Simple random sampling
Input data set	A
Random number seed	937359000
Sample size	10
Selection probability	0.5
Sampling weight	2
Output data set	B

Above is the basic information of the sampling. The sampling method is simple random sampling. Since the program does not assign the initial seed by using the option “seed=” for random number generation, the default seed is used as the initial seed (SEED=937359000). If the same sample in a subsequent execution is required, the same seed value in the “seed=” option should be specified. The expected sample size is 10. The notion “selection probability” refers to the probability of the sample being selected. Here the selection probability is 0.5. The selection probabilities of the complete random sampling are the same for each sample. “Sampling weight” refers to the reciprocal of the selection probability, which reflects how much information a sample contains.

The following table is the output data set generated by the procedure PRINT, which includes all the selected samples. The second column “ID” stands

for the numbers of the patients being selected.

Objects	ID	Gender	Age (years)
1	5	Female	41
2	6	Female	31
3	10	Male	16
4	11	Male	58
5	12	Male	63
6	13	Female	23
7	15	Female	20
8	16	Female	33
9	19	Female	49
10	20	Female	42

REFERENCES

- 1 Fang JQ. Medical statistics and computer experiments. 2nd ed. Shanghai: Shanghai Scientific and Technical Publishers. 2001. 199-201. Chinese.
方积乾. 医学统计学与电脑实验. 第 2 版. 上海: 上海科学技术出版社. 2001: 199-201.
- 2 Hu LP. Research design and statistical analysis of laboratory medicine. Beijing: People’s Military Medical Press. 2004: 4-8. Chinese.
胡良平. 检验医学科研设计与统计分析. 北京: 人民军医出版社. 2004: 4-8.

随机化分组的方法

胡良平, 鲍晓蕾, 王琪

军事医学科学院生物医学统计学咨询中心, 北京 100850

摘要: 随机化是试验设计 4 个基本原则之一。随机化通常分为从总体中随机抽取样品和将全部样品随机分组。前者简称为随机抽样, 而后者简称为随机分组。在随机分组中, 又可细分为完全随机分组、分层随机分组和动态随机分组。本文将介绍完全随机化的步骤、动态随机分组的概念以及用 SAS 软件实现随机抽样与随机分组。

关键词: 统计学, 医学; 研究设计; 随机分配; SAS 程序