

利用 SPSS 进行随机化实验设计分组

广州暨南大学医学院卫生统计教研室(510632) 林汉生 夏苏建

随机化(randomization)分组是指将受试对象按照随机的原则进行分组,是实验设计中保证非处理因素均衡的一个重要手段。只有通过随机分组,才能避免出现各种人为的客观因素和主观因素的偏性,提高统计检验效能。随机化分组也是统计检验假设用于推论因果关系的基础。然而,目前医学统计教科书一般只介绍用随机数字表或随机排列表进行分组。随着计算机的广泛运用,医学科研人员对 SPSS 统计软件的使用不断增加。为充分发挥统计软件的作用,本文介绍 SPSS 编程在随机分组中的应用。

一、完全随机设计分组 SPSS 编程

1. SPSS 的编程可以通过打开语句编辑窗口(Syntax Editor)得以实现。方法是通过菜单选择: File → New → Syntax。如果已经建立了程序,则可以通过 File → Open → Syntax 直接打开。

2. 在语句编辑窗口,用键盘输入程序,也可以在 Word 或者其他文本编辑软件中编辑以下程序,然后通过复制将程序粘贴到语句编辑窗口。程序中的英文字母不分大小写。

程序如下(注:右边的编号不必输入,只是便于后面的命令解释):

```
SET SEED= 12345. (1)
INPUT PROGRAM. (2)
LOOP NUMBER= 1 TO N. (3)
COMPUTE RANDOM= UNIFORM(N). (4)
END CASE. (5)
END LOOP. (6)
END FILE. (7)
END INPUT PROGRAM. (8)
AUTORECODE VARIABLES = RANDOM/ IN-
TO RANK. (9)
SORT CASES BY RANK(A). (10)
```

3. 程序的运行:在语句编辑窗口,通过菜单选择: Run → All 运行所有命令后,在数据编辑窗口就产生了完全随机设计分组结果。

4. 程序的解释:

第 1 句中的 SET SEED 是设定种子,取值在 1 到 200000 之间,其作用在于一旦设定后,每次运行得出同样结果。如果希望重复同样的分组结果,则可以设置该命令,否则可以省略该句。第 2~ 8 句的命令主要

是产生观察单位编号(等同于实验对象的编号),N 为待分组的实验对象总数。利用 COMPUTE RANDOM = UNIFORM(N) 命令产生随机数字(取值在 0~ N 之间)。如果将随机数字的小数点位数从默认的两位增加到 4 位或以上,可见随机数字无重复。第 9 句的作用是将产生的随机数字编秩后赋值给变量 rank。第 10 句是将 rank 变量值按照从小到大进行升序排列。

二、完全随机设计分组 SPSS 编程运用实例

某医生为了研究一种降血脂新药的临床疗效,按统一纳入标准选择了 120 名高血脂患者,采用完全随机设计方法将患者等分为 4 组进行随机临床对照试验。

首先将 120 名高血脂患者按就诊顺序从 1 到 120 编号。启动 SPSS 系统后,在 Syntax Editor 窗口内编写 SPSS 程序。注意:最后 1 句“EXECUTE”此处不能省略,否则倒数第 2 句不能执行。运行程序后,随机分组结果即显示在 Data Editor 窗口中。SPSS 程序编制如下:

```
INPUT PROGRAM.
LOOP NUMBER= 1 TO 120.
COMPUTE RANDOM= UNIFORM(120).
END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
AUTORECODE VARIABLES = RANDOM/ IN-
TO RANK.
SORT CASES BY RANK(A).
RECODE RANK(1 THRU 30= 1) (31 THRU 60
= 2) (61 THRU 90= 3) (91 THRU 120= 4) INTO
GROUP.
EXECUTE.
```

运行结果如下图 1 所示。

程序说明与结果解释:运行该程序后,产生观察单位编号(number)和随机数字(random),并将全部随机数字从小到大编序号(rank),按预先规定的序号 1~ 30 为第 1 组(group),序号 31~ 60 为第 2 组,序号 61~ 90 为第 3 组,序号 91~ 120 为第 4 组。结果如图所示,编号为 62, 96, 76, 73, 81, 117, 10, 85, 28, 79... 等的患者被纳入了第 1 组。由于在上述程序中没有设定

种子,即忽略了 SET SEED 命令,因此,每次运行上述程序均随机产生不同的分组结果。

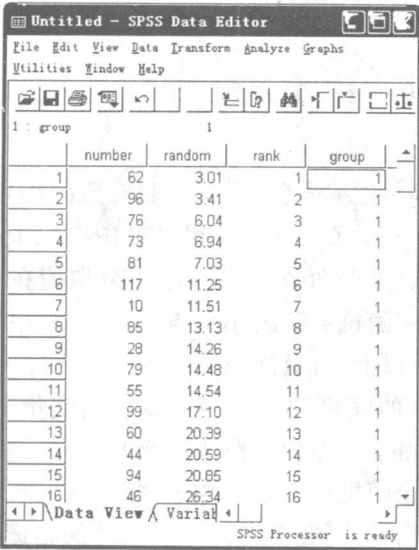


图 1 完全随机设计分组结果

方法为先将小白鼠的体重从轻到重编号,体重相近的 3 只小白鼠配成一个区组。将第 2 行和第 8 行的 N= 15,第 3 行的 K= 3,代入上述随机区组设计 SPSS 分组程序,然后运行之。

运行结果出现在数据编辑(Data Editor)窗口中,如图 2 所示。

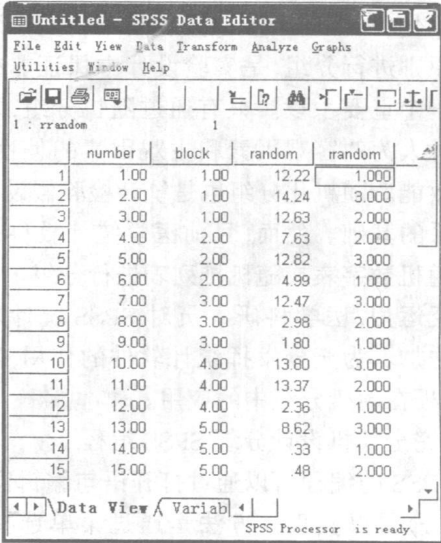


图 2 随机区组设计分组结果

number 为小白鼠的编号,block 为配伍组号,共有 5 组,每个配伍组有体重相近的 3 只小白鼠;random 是按从小到大顺序排列的随机数字;rrandom 是每个区组中随机数字的排列序号,序号范围是 1~ 3。按事先规定,序号为 1 的分到甲组,序号为 2 分到乙组,序号为 3 分到丙组。可将 SPSS 分组结果整理成下表。

表 1 5 个区组小白鼠按随机区组设计分组结果

| 区组 | 甲 | 乙 | 丙 |
|----|----|----|----|
| 1 | 1 | 3 | 2 |
| 2 | 6 | 4 | 5 |
| 3 | 9 | 7 | 8 |
| 4 | 12 | 11 | 10 |
| 5 | 14 | 15 | 13 |

完全随机化设计、随机区组设计是实验设计和临床试验设计中广泛采用的方法。传统上常用的随机化分组方法有随机数字表法和随机排列法,虽简单易行,但不适用于大样本,如在完成大规模的多中心临床试验时操作繁琐、工作效率低下,并且容易出错。运用 SPSS 编程可以轻松自如地解决上述随机分组中的问题,具有很强的可操作性,可提高工作效率,减少误差。

三、随机区组设计分组的 SPSS 编程
通用的随机区组设计分组 SPSS 程序如下(右边的编号不必输入):

```
INPUT PROGRAM. (1)
LOOP NUMBER= 1 TO N. (2)
  COMPUTE BLOCK = RND((NUMBER - 1) / K
+ 0. 5). (3)
END CASE. (4)
END LOOP. (5)
END FILE. (6)
END INPUT PROGRAM. (7)
COMPUTE RANDOM = UNIFORM(N). (8)
RANK VARIABLES= RANDOM BY BLOCK. (9)
```

上述程序中,N 为观察单位总数,K 为处理组数。第(2)语句产生 1~ N 的观察单位编号,number;第(3)语句产生观察单位对应的区组编号 block(1~ K);第(8)语句产生随机数字 random(取值在 0~ N 之间);第(9)语句是以区组 block 为分组变量,将随机数字 random 编秩,并自动赋值给新变量 rrandom。当 N 为偶数,处理组数 K= 2 时,即为配对设计。

四、随机区组设计分组 SPSS 编程运用实例
如何按随机区组设计,分配 5 个区组的 15 只小白鼠接受甲、乙、丙三种抗癌药物处理?