

Using PROC PLAN for Randomization Assignments

Miriam W. Rosenblatt

Division of General Internal Medicine and Health Care Research, University
Hospitals of Cleveland

Abstract

This tutorial is an introduction to using PROC PLAN, and includes examples of randomization. Although PROC PLAN is not an easy procedure to master, it is extremely useful for doing random assignments. Data handling ideas for using this procedure, such as combining PROC FORMAT with PROC PLAN, allow the user to create formatted reports of random assignment. A user-friendly report can then be used for the preparation of randomization envelopes, thus ensuring that a given randomization plan is implemented accurately.

Introduction

PROC PLAN is a valuable SAS® procedure that constructs randomization plans for all kinds of experiments. The randomization can be a simple run of random numbers or a more sophisticated experimental design. The SAS/STAT documentation presents PROC PLAN for use with sophisticated randomization designs, such as nested, hierarchical, and Latin Square designs. However, PROC PLAN can easily be used for more basic randomization designs, without having to use the data step or direct manipulation of data.

This tutorial provides an introduction to using PROC PLAN for basic types of randomization, and shows how to use the results to create user-friendly reports for the implementation of the randomization. I will provide simple examples to illustrate the procedures, as well as tips and tricks I have used in doing randomization.

Why Randomize?

The main purpose of randomization is to select a study sample that represents the population to be studied, thus allowing for generalization of final results. Moreover, the sample allows the researcher to determine any measurement errors based on the given estimate. Randomization is also used to assign subjects into treatment groups so that subjects have an equal chance of being chosen, avoiding any selection bias in each of the final study samples.

Randomization Basics

PROC PLAN generates a list of random numbers based on a uniform random distribution. A seed number can be supplied to start the random number generator for selecting factor levels randomly. A seed number (any positive integer up to $2^{31} - 1$) can be supplied by the user. If a seed number is not given, then SAS will use the time of the day, based on the computer clock. Because using the default may result in generating artificial correlations, it is recommended that the user supply the seed number. I frequently use a date value for a seed number, which will assure that the run is unique.

When generating a final randomization list, I like using PROC FORMAT along with PROC PLAN. PROC PLAN can create a SAS dataset that can be used to generate a report of the randomization scheme. PROC FORMAT can be used to associate a range of numbers generated by PROC PLAN for the creation of user-friendly reports. These final reports can more easily be understood by the user for the selection and implementation of random assignments.

The basic mechanics of generating a user-friendly randomization list are outlined below, along with a sketch of the SAS code.

1. Format Statements

Create a format statement with a number string you plan to use. This will be used with the final randomization report.

```
PROC FORMAT;  
VALUE GROUP 1, 3, 4, 7, 9 = 'Treatment A'  
            2, 5, 6, 8, 10 = 'Treatment B';
```

2. PROC PLAN

Use PROC PLAN to generate the randomization list.

```
PROC PLAN SEED=020895;  
  OUTPUT OUT=PLANDAT1;  
  FACTORS UNIT=1 RANDOM GROUP = 10;  
  TITLE 'RANDOMIZATION ASSIGNMENT: FOR 10 SUBJECTS';
```

3. Print out the report with the format statements.

```
PROC PRINT DATA=PLANDAT1 D N;  
  FORMAT GROUP GROUPF.;
```

Most of the examples below will illustrate the use of formats with randomization reports.

4. Implement the randomization assignments

The final step of generating the randomization report is implementation. Having a user-friendly report will facilitate implementation of a randomization scheme, avoiding misinterpretation of the results, and ensuring that whoever implements the final randomization can be blinded to a given randomization scheme. Prepare a set of envelopes and a corresponding set of insert forms, numbering each envelope and insert set with a number. Make sure the number is also placed on the original random list for reference. Each form should contain at least the following information: randomization number, treatment, final randomization status of the subject (enrolled, not enrolled, reason not enrolled, and any protocol violations). The insert can also be designed as a data entry form, and entered into a spreadsheet or a data file. This is useful for prospective tracking and summarization of the randomization process.

Randomization Examples

1. Simple Randomization

Example: You have a mailing list of 25 people, and you want to sample the first 10 people to mail them a survey. To do this you would create a random string of 25 numbers and take the top 10 subjects from the list. The report is located in **Appendix 1, OUTPUT 1.**

```
PROC PLAN SEED=123123;  
  OUTPUT OUT=EX1;  
  FACTORS UNIT=25 RANDOM;  
  TITLE 'EXAMPLE 1';  
  TITLE2 'SIMPLE RANDOM STRING OF 25';  
PROC PRINT;
```

2. Assignment to Two Treatments

Example: You want to assign 20 subjects to either treatment A or the control treatment. You have decided that an odd number will be assigned to A, and the even numbers to the placebo group.

```
PROC PLAN SEED=123567;  
  OUTPUT OUT=EX2;  
  FACTORS UNIT=50 RANDOM;  
  TITLE2 'EXAMPLE 2: TWO TREATMENT ALLOCATIONS';  
PROC PRINT;
```

To make the output more readable, use PROC FORMAT.

```
PROC FORMAT;  
  VALUE TREATF 1, 3, 5, 7, 9, 11, 13, 15, 17, 19 = 'TREATMENT A'  
              2, 4, 6, 8, 10, 12, 14, 16, 18, 20 = ' PLACEBO';  
  
PROC PRINT D N;  
  FORMAT UNIT TREATF.;
```

The results are shown in **OUTPUT 2**.

Stratification of Two Treatments

For some studies, you may be interested in ensuring that appropriate subgroups are assigned to two treatments in equal numbers, and that each subgroup is not under- or over-sampled. For example, you are interested in people who are 60 and older and want to make sure you have equal numbers in each treatment group for your study. Subjects are selected randomly from each subgroup or stratum into which they fall. One subgroup would include people who are 60 and older (Set A). The other subgroup would include people who are under the age of 60 (Set B). To do this, PROC PLAN would be run two times to generate a random number string for each group. This would finally result in two sets of "envelopes", one to be used for each age group, depending on the age of the given subject entering the study. Therefore if a 40-year-old eligible subject were to be randomized, the first envelope from "Set A" would be opened and the treatment assigned on the basis of its contents.

3. Blocked Design

In randomization, blocking is used to assure equal sample sizes within a fixed group size. In the example below, there would be equal sample sizes between Treatment A and the placebo for every group of 4 subjects. When implementing this kind of randomization, it may important to make sure that people who assign the randomization are blinded to the block size, to ensure that they cannot "predict" assignment.

```
PROC PLAN SEED=123567;  
  OUTPUT OUT=EX4;  
  FACTORS UNIT=25 RANDOM GROUP = 4 RANDOM;  
  TITLE2 'EXAMPLE 4: BLOCKED STUDY DESIGN';  
PROC FORMAT;  
  VALUE TREATF 1, 3 = 'TREATMENT A' 2, 4 = ' PLACEBO';  
PROC PRINT;  
  FORMAT GROUP TREATF.;
```

The results are shown in **Output 3**.

4. Using Proc Plan To Randomize from a SAS Dataset

Some of my randomization applications involve sampling from a SAS dataset. For example, I have a SAS dataset of 200 subjects, and I want to randomly sample 10%. First, I would use PROC PLAN to generate a SAS dataset containing the random list of 20, giving the generated random list the same variable name as the subject ID of the original SAS dataset. The random file is sorted by subject number, and match-merged with the source data by subject ID keeping only the selected subjects (using the IN= option in the merge statement). This is especially useful in generating a sample from a mailing list.

```
PROC PLAN SEED=123567;
  OUTPUT OUT=SUBJLIST;
  FACTORS SUBJECT=200 RANDOM;
  TITLE2 'EXAMPLE 4: RANDOM SUBJECT LIST';
DATA SAMPLE; SET SUBJLIST(OBS=20);
PROC PRINT DATA=SAMPLE;
TITLE3 'SELECT THE FIRST 20'; PROC SORT DATA=SAMPLE; BY SUBJECT;
DATA SELECT;
  MERGE SUBJLIST (IN=INSAMPLE)
        MYLIB.MYDATA
  ; BY SUBJECT; IF INSAMPLE;
PROC PRINT;
  VAR SUBJECT LNAME FNAME;
  TITLE3 'RANDOM LIST SELECTED FROM A SAS DATASET';
PROC PRINT DATA=SELECT (OBS=3);
  VAR SUBJECT LNAME FNAME;
  TITLE3 'FINAL SELECTION';
```

The results are shown in **Output 4**.

Discussion

PROC PLAN is a valuable procedure that is not just for sophisticated randomization designs. PROC PLAN may have some advantages over using data steps randomize, especially for doing blocked or stratified randomization designs. This procedure can be used to generate a user-friendly report, facilitate implementing a randomization scheme, and assuring that the people executing random assignment are blinded to the number scheme, where appropriate. Random assignment can be done using the RANUNI function in the data step, which can involve data processing. I switched to using PROC PLAN when I began doing blocked and stratified randomizations, and hope others find it as useful as I do as an alternative to data processing.

References:

1. SAS Institute, Inc. (1990), SAS Language Reference, Version 6, First Edition, Cary, NC: SAS Institute Inc.
2. SAS Institute, Inc. (1990), SAS Procedures Guide, Version 6, Third Edition, Cary, NC: SAS Institute Inc.
3. SAS Institute, Inc. (1990), SAS/STAT User's Guide, Version 6, Fourth Edition, Cary, NC: SAS Institute Inc.
4. B.C. Decker, Inc. (1989), PDQ Epidemiology, Streiner, Norman, Blum

SAS, SAS/ ACCESS are registered trademarks of SAS Institute, Inc. in the USA and other countries. ® indicates USA registration. Other brands and product names are registered trademarks or trademarks of their respective companies.

I would like to thank Barbara Juknialis for her editorial advice and Linda M. Quinn for her review of the manuscript.

Miriam W. Rosenblatt
 Division of General Internal Medicine and
 Health Care Research
 University Hospitals of Cleveland
 11100 Euclid Avenue
 Cleveland, Ohio 44106

Appendix 1:

SAS Programming and OUTPUT

Output 1: Simple Randomization

EXAMPLE 1

Procedure PLAN

Factor	Select	Levels	Order
UNIT	25	25	Random

UNIT

4	7	16	24	14	12	11	3	22	23	25	19	6	2	9
20	1	13	8	21	17	18	5	15	10					

25 NUMBERS IN RANDOM ORDER

OBS UNIT

1	4
2	7
3	16
4	24
5	14
6	12
7	11
8	3
9	22
10	23
11	25
12	19
13	6
14	2
15	9
16	20
17	1
18	13
19	8
20	21
21	17
22	18
23	5
24	15
25	10

OUTPUT 2 : Treatment Allocations

EXAMPLE 2: 2 TREATMENT ALLOCATIONS

Procedure PLAN

Factor	Select	Levels	Order
UNIT	20	20	Random

UNIT

6	20	4	5	15	16	11	3	19	14	18	12	8	9	7
---	----	---	---	----	----	----	---	----	----	----	----	---	---	---

17 2 10 1 13

PROC PLAN

EXAMPLE 2: 2 TREATMENT ALLOCATIONS

OBS UNIT

1	6
2	20
3	4
4	5
5	15
6	16
7	11
8	3
9	19
10	14
11	18
12	12
13	8
14	9
15	7
16	17
17	2
18	10
19	1
20	13

ODD NUMBERS: TREATMENT A
 EVEN NUMBERS: PLACEBO

PROC PLAN

EXAMPLE 2: 2 TREATMENT ALLOCATIONS

FORMATTED OUTPUT

OBS UNIT

1	PLACEBO
2	PLACEBO
3	PLACEBO
4	TREATMENT A
5	TREATMENT A
6	PLACEBO
7	TREATMENT A
8	TREATMENT A
9	TREATMENT A
10	PLACEBO
11	PLACEBO
12	PLACEBO
13	PLACEBO
14	TREATMENT A
15	TREATMENT A
16	TREATMENT A
17	PLACEBO
18	PLACEBO
19	TREATMENT A
20	TREATMENT A

N = 20

ODD NUMBERS: TREATMENT A
 EVEN NUMBERS: PLACEBO

OUTPUT 3:

Blocked Design

PROC PLAN
EXAMPLE 3: 25 UNITS OF
UNIT 4

Procedure PLAN

Factor	Select	Levels	Order
UNIT	25	25	
Random			
GROUP	4	4	Random

UNIT	GROUP				
22	4	3	2	1	
7	4	1	3	2	
19	2	1	3	4	
24	1	2	3	4	
23	4	3	2	1	
21	3	2	4	1	
17	4	2	1	3	
10	4	3	1	2	
13	1	4	3	2	
8	4	2	3	1	
9	1	4	2	3	
6	2	4	3	1	
18	2	4	3	1	
1	3	2	4	1	
2	3	1	4	2	
14	2	3	1	4	
20	2	4	1	3	
4	2	1	3	4	
3	4	1	3	2	
15	2	4	3	1	
12	1	4	3	2	
5	1	3	4	2	
11	4	3	2	1	
25	2	4	1	3	
16	2	4	3	1	

ODD NUMBERS:
TREATMENT A
EVEN NUMBERS:
PLACEBO

PROC PLAN
EXAMPLE 3: 25 UNITS OF
UNIT 4

OBS UNIT GROUP

1	22	4
2	22	3
3	22	2
4	22	1
5	7	4
6	7	1
7	7	3
8	7	2
9	19	2
10	19	1
11	19	3
12	19	4
13	24	1

14	24	2
15	24	3
16	24	4
17	23	4
18	23	3
19	23	2
20	23	1
21	21	3
22	21	2
23	21	4
24	21	1
25	17	4
26	17	2
27	17	1
28	17	3
29	10	4
30	10	3
31	10	1
32	10	2
33	13	1
34	13	4
35	13	3
36	13	2
37	8	4
38	8	2
39	8	3
40	8	1
41	9	1
42	9	4
43	9	2
44	9	3
45	6	2
46	6	4
47	6	3
48	6	1
49	18	2
50	18	4

ODD NUMBERS:
TREATMENT A
EVEN NUMBERS:
PLACEBO

PROC PLAN
EXAMPLE 3: 25 UNITS OF
UNIT 4
FORMATTED OUTPUT

OBS	UNIT	GROUP
1	22	PLACEBO
2	22	TREATMENT A
3	22	PLACEBO
4	22	TREATMENT A
5	7	PLACEBO
6	7	TREATMENT A
7	7	TREATMENT A
8	7	PLACEBO
9	19	PLACEBO
10	19	TREATMENT A
11	19	TREATMENT A
12	19	PLACEBO
13	24	TREATMENT A
14	24	PLACEBO
15	24	TREATMENT A
16	24	PLACEBO
17	23	PLACEBO
18	23	TREATMENT A

19	23	PLACEBO
20	23	TREATMENT A
21	21	TREATMENT A
22	21	PLACEBO
23	21	PLACEBO
24	21	TREATMENT A
25	17	PLACEBO
26	17	PLACEBO
27	17	TREATMENT A
28	17	TREATMENT A
29	10	PLACEBO
30	10	TREATMENT A
31	10	TREATMENT A
32	10	PLACEBO
33	13	TREATMENT A
34	13	PLACEBO
35	13	TREATMENT A
36	13	PLACEBO
37	8	PLACEBO
38	8	PLACEBO
39	8	TREATMENT A
40	8	TREATMENT A
41	9	TREATMENT A
42	9	PLACEBO
43	9	PLACEBO
44	9	TREATMENT A
45	6	PLACEBO
46	6	PLACEBO
47	6	TREATMENT A
48	6	TREATMENT A
49	18	PLACEBO
50	18	PLACEBO

100 16 TREATMENT A

N=100

OUTPUT 4:

Randomizing from a SAS Dataset

EXAMPLE 4:
RANDOMIZE 200

Procedure PLAN

Factor	Select	Levels	Order
SUBJECT	200	200	
Random			

SUBJECT					
59	159	161	164	93	2
162	147	75	98	150	138
196	101	145	30	184	169
171	125	168	181	17	25
83	186	103	50	42	139
47	124	120	118	56	49
157	88	3	68	67	57
29	129	92	198	100	193
137	54	85	149	176	182
80	113	60	82	70	108
152	34	84	199	141	58
151	64	4	99	185	111

48	36	16	46	117	66	192
187	114	40	62	45	133	
69	31	126	63	20	38	189
23	122	89	156	195	90	
190	200	175	191	116	188	
24	180	148	65	28	136	
95	10	166	6	79	91	87
142	94	134	178	18	9	
76	35	174	119	78	27	
173	55	102	74	107	96	
21	170	179	160	130	1	
81	7	86	33	110	5	43
172	132	112	52	12	154	
22	14	197	53	158	44	41
106	39	140	109	77	13	
123	143	115	104	127	165	
128	131	177	15	121	155	
71	146	37	194	73	61	
167	163	183	135	51	26	
8	97	32	153	11	72	19
144	105					

OBS SUBJECT

1	59
2	159
3	161
4	164
5	93
6	2
7	162
8	147
9	75
10	98
11	150
12	138
13	196
14	101
15	145
16	30
17	184
18	169
19	171
20	125

PROC PLAN

EXAMPLE 4: FINAL
SELECTION

OBS SUBJECT LNAME

1	59	MALLET
2	159	KARAS
3	161	KOVAL

N = 3