Solutions to Selected Problems from Homework 1

5. The goal is to develop blocks of experimental units so that the experimental units within a block are similar with respect to characteristics that might be associated with the response variable. The only characteristics of the experimental units that are known to us are age and gender. Both could be associated with respiratory and cardiovascular measurements. Thus we want to create blocks of experimental units where units within the same block are similar with respect to age and gender. We can begin by sorting the experimental units according to age and gender.

Individual	Sex	Age	Individual	Sex	Age
4	F	18	10	M	18
12	F	18	11	M	31
5	F	19	18	M	34
16	F	21	17	M	35
19	F	38	2	M	38
6	F	39	7	M	51
3	F	41	1	M	54
8	F	44	13	M	58
15	F	58	9	M	62
			14	M	74

It looks as if the youngest four females and the oldest five females will make two nice blocks containing experimental units that are relatively homogeneous with respect to age (and completely homogeneous with respect to gender). Separating the two groups of males into the youngest 5 and the oldest 5 seems natural because there is a large difference in age between the 38 year-old subject 2 and the 51 year-old subject 7 males. Initial groups of homogeneous subjects are as follows.

Group 1			**Group 2**			**Group 3**			**Group 4**		
Individual	Sex	Age	Individual	Sex	Age	Individual	Sex	Age	Individual	Sex	Age
4	F	18	19	F	38	10	M	18	7	M	51
12	F	18	6	F	39	11	M	31	1	M	54
5	F	19	3	F	41	18	M	34	13	M	58
16	F	21	8	F	44	17	M	35	9	M	62
			15	F	58	2	M	38	14	M	74

We only require 16 experimental units, so it will be nice to obtain blocks of size four. Group 1 is already a block of size 4. The other groups can be limited to 4 experimental units each by randomly eliminating one member of each group. Alternatively we could exclude the oldest members in groups 2 and 4 and the youngest member in group 3. These individuals have ages that are quite different from the rest of the individuals in their groups. If this latter strategy is adopted, it must be made clear that some individuals have been excluded because they were quite a bit older or quite a bit younger than other volunteers of their gender. This can limit the ability to generalize the results of the experiment to broad age ranges but will likely result in lower estimates of error variances. Because it is slightly more complicated, we will illustrate the other strategy of randomly selecting four units from each group. Starting with line 6 on page 625 of Table XXII, we assign each group member a random digit 0 through 9 in such a way that no repeats are allowed within a group. (If a digit is selected that has already been chosen for another subject in the group, we skip that digit and try the next one.)

Group 1

Individual	Sex	Age	Random Digit
4	F	18	6
12	F	18	9
5	F	19	5
16	F	21	3

Group 2

Individual	Sex	Age	Random Digit
19	F	38	0
6	F	39	1
3	F	41	7
8	F	44	5
15	F	58	3

Group 3

Individual	Sex	Age	Random Digit
10	M	18	5
11	M	31	2
18	M	34	7
17	M	35	8
2	M	38	3

Group 4

Individual	Sex	Age	Random Digit
7	M	51	8
1	M	54	4
13	M	58	0
9	M	62	6
14	M	74	1

Experimental units are arranged within groups from smallest random digit to largest. The first two observations in each group are assigned treatment A. The next two are assigned treatment B. Any remaining experimental units are discarded.

Group 1

Individual	Sex	Age	Random Digit	Treatment
16	F	21	3	A
5	F	19	5	A
4	F	18	6	B
12	F	18	9	B

Group 2

Individual	Sex	Age	Random Digit	Treatment
19	F	38	0	A
6	F	39	1	A
15	F	58	3	B
8	F	44	5	B
3	F	41	7	-

Group 3

Individual	Sex	Age	Random Digit	Treatment
11	M	31	2	A
2	M	38	3	A
10	M	18	5	B
18	M	34	7	B
17	M	35	8	-

Group 4

Individual	Sex	Age	Random Digit	Treatment
13	M	58	0	A
14	M	74	1	A
1	M	54	4	B
9	M	62	6	B
7	M	51	8	-

The end result is a randomized complete block design. Block 1 consists of individuals 16, 5, 4, and 12. Block 2 consists of individuals 19, 6, 15, and 8. Block 3 consists of individuals 11, 2, 10, and 18. Block 4 consists of individuals 13, 14, 1, and 9. The disadvantage of this strategy is best seen by examining the assignment of treatments to experimental units in group 2. The females receiving treatment B are noticeably older than the females receiving treatment A. This could bias the results if age plays a large role in the response of interest. We could force a more equal random assignment by making the blocks even smaller (e.g., use blocks with 2 experimental units instead of blocks with 4 experimental units).

6. (a) Method 2 is the best. (b) This is a randomized complete block design. A block is a set of three shirts – with one shirt from each treatment – that are placed together in the simulation machine for a single run. The treatments are applied individually to each shirt so that each shirt is an experimental unit. There are 4 replications per treatment, one in each block. It will be possible to make comparisons of the treatments within each block so that the effect of different runs of the simulation machine won't bias our comparisons of the treatments. (c) Method 1 is bad for two main reasons. First the experimental unit is a batch of four shirts because each treatment is applied simultaneously to a batch of four shirts. There is no way to estimate error variance because there is only one replication per treatment. Second there is no way to separate the effect of the run of the simulation machine from the effect of the treatment. Any observed differences could be due to differences in runs and/or differences in treatments. Treatment and run of the simulation machine are said to be *confounded* because we can't separate their effects on the response. Even if the shirts were treated

individually as in Method 3, the confounding of treatment with the run of the simulation machine would remain.

8. (a) Treatments are applied to instructor/class combinations. Thus instructor/class combination is the experimental unit. (Depending on what response variable will be measured, students could be the observational units. Many times students are incorrectly assumed to be the experimental units in education-method experiments.) We have only one experimental unit per treatment. Thus there is no way to measure experimental error and no way to determine if observed differences in the response are statistically significant. Even if we believed observed differences were statistically significant, there would be no way to know if the differences were due to classes, teachers, or the teaching methods. All are confounded with one another. (b) Using a completely randomized design, the experiment would require several instructor/class combinations. Each of the three methods would be assigned at random to 1/3 of the instructor/class combinations. An hypothetical example with 9 classes and 9 instructors is provided below.

Class	Instructor	Method
1	1	B
2	2	C
3	3	C
4	4	A
5	5	B
6	6	B
7	7	A
8	8	A
9	9	C

Most likely more instructor class combinations are unavailable. It might be conceivable to use a randomized complete block design if each class was split into 3 classes randomly. Each instructor would teach a group using each of the methods. An example design is provided below. There are 9 classes in the example because we are assuming that each of the original 3 classes has been split into 3 classes. The classes would randomly be assigned to instructor and the methods of teaching assigned at random to the three classes taught by each instructor.

Class	Instructor	Method
7	1	B
2	1	A
6	1	C
9	2	A
5	2	B
8	2	C
1	3	A
3	3	C
4	3	B

Classes are experimental units in this case. Three classes taught by the same instructor form a block. Each treatment appears exactly once in each block.

You might suggest that the same teacher be used to teach all classes. However, with only three classes we would still have the problem of no replication. We cannot think of the students as experimental units even if they are randomly assigned to classes because the treatments are not *independently* applied to the students. There would be confounding with factors such as time of day or day of the week because the three classes could not possibly be taught at the exact same time. The students may be able to learn better and/or the teacher may be able to teach better at certain times of the day or week. Thus using one teacher for all three classes does not solve the problem. Multiple classes need to be taught with each method in order to avoid confounding.

Even with replication of classes, inability to generalize the results would be another problem with using only one instructor. If method A is the best for the instructor used in the experiment, should we assume that method

A is the best for all instructors?

9. (a) The observational units are asphalt specimens. The experimental units are batches. Thus there is only one replication per treatment, and we have no way of measuring experimental error. Observed differences between the treatments could be due to differences in batch preparation. To illustrate the problems that arise, imagine that the experiment was repeated as described except that only one experimental mixture was used for the preparation of each batch. In other words, assume the three treatments are really one treatment repeated three times. We would not get the exact same tensile strength measurements for the specimens in the batches. One batch would certainly have the highest mean. It would be nonsense to think there is a difference in treatments because each batch received the exact same treatment. We would only be observing differences in batches. When the experiment is conducted as described we have no way of knowing if we are observing differences in batches, differences in experimental mixtures (i.e., treatments), or both. (b) Use multiple batches for each experimental mixture.

10. There are 16 choose 8 ways to randomly assign the experimental units into two groups of 8.

    16 choose 8 = 16!/(8!*(16-8)!) = 12870

    There are 16 choose 6 ways to randomly assign the experimental units into one group of 6 and one group of 10.

    16 choose 6 = 16!/(6!*(16-6)!) = 8008.

12. There are 6 choose 3 ways to randomly assign labels A and B equally to the observations. Thus, there are 6!/(3!*3!) = 20 possible arrangements. Below are half of these possible arrangements. The other half may be obtained by taking the ”mirror” of these 10.

Arrange- Ment	Unit: Response:	1 5	2 7	3 9	4 10	5 10	6 12	Mean B - Mean A
1		A	A	A	B	B	B	3.67
2		A	A	B	A	B	B	3
3		A	A	B	B	A	B	3
4		A	A	B	B	B	A	1.67
5		A	B	A	A	B	B	1.67
6		A	B	A	B	A	B	1.67
7		A	B	B	A	A	B	1
8		A	B	A	B	B	A	0.33
9		B	A	A	A	B	B	0.33
10		B	A	A	B	A	B	0.33

Arrangement #2 with a mean difference of 3 is the arrangement observed in the experiment. The alternative hypothesis is a one-way test, so we are only interested in the number of arrangements that result in a mean B - mean A difference of 3 or more. The first three arrangements listed meet this criterion and we can see that all the “mirror” arrangements will not meet this criterion. Thus, the one-sided p-value for our randomization test is 3/20 = 0.15. Because this p-value is somewhat large, we cannot rule out the possibility that observed difference resulted from chance assignment of experimental units to groups rather than a treatment effect. Therefore we fail to reject the null hypothesis that there is no difference between treatments. It may be that the effect of treatment B is no different than the effect of treatment A.