

Weeks 8 and 9: Lecture – Randomized Block (RBD)

Experimental designs beyond the completely randomized design can be classified into two types. Using terminology from signal detection theory. We can call one class of designs as “volume-increasing” designs and the other as “noise-reducing” designs. In this module we will discuss one of the noise-reducing designs. This design is called *randomized blocks*. Here in this design there is still only one independent variable or set of treatment conditions. The difference between the completely randomized design and the randomized block design lies in the selection of participants.

A fundamental principle for reducing the noise (error) of an experiment is to use homogeneous experimental units. The more similar or homogeneous, the easier it is to find treatment differences. In the physical sciences, and other more exact scientific areas, the homogeneous material used in experiments is so cultivated that statistical methods of analysis are infrequently used because the amount of experimental error is so small. In most behavioral and social science research, the situation is different. The behavioral scientist must work with relatively heterogeneous material, and experimental error plays a fundamental role in data analysis. In the behavioral sciences homogeneous experimental units might take the form of using participants that are highly similar to one another so that only a small part of their responses is a product of individual characteristics and the bulk being attributed to the effects of the treatments. In some cases, the same subject is used. Since noise-reducing designs work with blocks of homogeneous experimental units they are referred to as block designs. The *randomized blocks* design provides some control over the diversity of experimental subjects. In these designs, one or more characteristics of the subjects that are known to be related to performance in the experimental situation are explicitly taken into account in the structure of the design. In effect, these variables operate in the same capacity as treatment variables, although the experimenter does not directly manipulate them.

Structurally, randomized blocks designs consists of two or more dimensions, and all possible combinations of these dimensions are applied to independent, random groups of experimental subjects. One or more of the dimensions represent classification variables used to group subjects into relatively homogeneous subclasses. In some experiments, the classification dimensions, or *blocking variables*, are introduced solely for the purpose of increasing the precision of the experiment relative to the manipulated treatment dimensions. In other cases, these blocking variables may be of intrinsic interest. For example, in an experiment concerned with classroom learning, the researcher may block students on the basis of previous academic performance since new learning will be related to previous performance. The researcher may not, however, be at all interested in studying this relationship per se, since it is well established. The aim is to introduce homogeneous subgroups in order to increase the precision of comparisons involving manipulated treatment variables. On the other hand, in the same experiment, the researcher may block students by sex; in this instance, the concern may be with whether or not performance is different between the two sexes and whether or not sex interacts with other treatment dimensions.

Later in the course when we study factorial designs, we will see that the randomized block design is structurally a factorial design with one or more of the classification variables. Some researchers have also referred to the randomized block design as a *levels-by-treatment* design.

It is important to clarify the manner in which the introduction of one or more blocking dimensions may increase the precision of an experiment. Consider the simplest experimental design, the completely randomized design. In its basic form, the variability of scores is due either to effects from the treatment levels or to experimental error. Within any one treatment group, the only source of variability among scores is experimental error. Of course, the smaller the amount of experimental error, the more sensitive will be the test for treatment effects, since MS.error will be

estimating a smaller variance. If a blocking dimension is added to the CRD, there will be formed, within each treatment group and across all treatment groups, relatively homogeneous subclasses of experimental subjects. The sum of squares associated with the blocking variable is removed directly from experimental error, and the treatment effects are unchanged. Thus, the MS_{error} is smaller, and the reduction in the amount of experimental error is related to the how much the blocking classifications do, in fact, result in homogeneous subclasses. That is, the more homogeneous, the greater the reduction in MS_{error} .

In the CRD design, the error mean square is computed by pooling the variances from the separate treatment groups. Assume that the subjects in the treatment groups differ in terms of some outside variable that is related to performance in the experiment (e.g., intelligence, anxiety, or sociability). Since subjects are randomly assigned to the treatment groups, this outside variable is randomly distributed over the groups. Thus, some of the variability *within* groups is due to differences in this outside variable. In the CRD that variability is unaccounted for in any explicit sense and merely becomes part of the within-cell-error variance. In the randomized block design, the variability due to the outside measure is taken into consideration. In fact we specify it as a known source of variability. Hence, that variability was removed from the within groups variance, leaving the resulting within group variance (i.e. MS_{error}) a smaller value. The total variability of scores in the design is the same as for the CRD. The partitioning of the total variability is different. That is, the total variability is split among treatment, blocks, and error, whereas in the CRD it is split between only treatment and error. Of course, more than one blocking dimension can be added to an experimental design. From a design point of view, the researcher must search for blocking variables which will, in fact, result in more homogeneous subgroups than would occur through the use of completely randomized groups. This consideration is mitigated somewhat by the fact that in some applications the researcher may introduce a blocking variable not so much to reduce error variability but because there is an interest in the effects of this blocking variable. In this event, the design is essentially no different from a factorial design except that the blocking dimension is a selection rather than a manipulated variable.

When the intent is to reduce error variability, the choice of blocking variables and the decision concerning how to establish the levels of such variables become critical. Often a variable that a researcher wishes to utilize as a blocking variable is continuous and not naturally occurring in discrete classes. Hence, arbitrary cutting points must be established to form the levels for a randomized blocks design. This could cause difficulties.

The major considerations in the choice of blocking variables are their relationship to the criterion variable and their availability prior to the formation of the experimental treatment groups. Randomization in a randomized blocks design must involve random assignment of subjects within each block to the various treatment levels, or combinations of treatment levels. Typically, the design requires that the subjects in block 1 be randomly assigned to the levels of the treatment and, similarly, that the subjects in block 2 be randomly assigned to the levels of the treatment and so on. This can be accomplished only if the blocks are formed prior to actually setting up the experimental conditions required for application of the treatment variable. Thus, the measurements that form the basis for blocking subjects must be taken before the experiment is executed. In education, this may be done through setting up special testing sessions, or it may only require going to available school records (e.g., prior achievement scores, intelligence scores, et cetera. These frequently used blocking variables are generally available from student cumulative records). The choice of a blocking variable should be made on the basis of its predictive value with respect to the criterion variable. In general, the greater the degree of correlation between the blocking variable and the criterion variable, the greater will be the increase in power of the randomized blocks design relative

to the completely randomized design. The direction of relationship is irrelevant; that is, a positive or a negative correlation can typify a satisfactory blocking variable. Thus, the researcher should select a blocking variable which is most likely to bear a close relationship to the criterion variable. An obvious choice is the use of a premeasure. This premeasure can be the measure of the criterion variable itself or some closely parallel form of that measure. For example, in a classroom learning study, a pretest may be given to all subjects; after blocks are formed and the experimental treatments applied, the same test may be used as the measure of the criterion. In most practical situations, the use of a blocking variable that is highly similar to or identical with the measure of the criterion is a simple way of assuring a high degree of correlation between the blocking and criterion variables. Ordinarily, the researcher's knowledge concerning the criterion measure will suggest one or more relevant blocking variables. If the experiment is focused on some aspect of academic performance involving achievement, the researcher can select or build a relevant aptitude or preachievement instrument; measures of general intelligence are commonly used, although a measure of a specific and relevant aptitude is generally a better predictor of academic criteria.

The word "randomization" in the randomized block design implies that the treatment conditions are randomly assigned within a block. The diagram below shows the blocking, treatment and participants.

| Participants | Blocks | | | |
|--------------|--------|---|---|---|
| | 1 | 2 | 3 | 4 |
| 1 | B | A | D | C |
| 2 | D | B | C | A |
| 3 | A | C | B | D |
| 4 | A | D | C | B |
| 5 | C | A | B | D |
| 6 | D | B | A | C |

Note that each participant receives all four-treatment conditions. They get the treatments in a different order. In fact the treatments were presented to each participant in random order. The capital letters (e.g., A, B, C and D) designate the treatments. With this design, the participant-to-participant variation can be eliminated.

If we state that q is the number of levels of the blocking variable, it is apparent that the loss of degrees of freedom from error is linearly related to q . Choosing a very large q in order to assure highly homogeneous levels of the blocking dimension comes at a price. If q is large, the analyses will reduce the degrees of freedom for error. Although not immediately apparent, there is an optimal number of levels of the blocking dimension for various sample sizes and various numbers of levels of A, the treatment variable. Feldt (1958) has tabled some of these optimal conditions. The apparent imprecision of a randomized blocks design is defined as the ratio of the sampling error averaged over blocks to the theoretically minimum sampling error; this ratio is adjusted in terms of degrees of freedom. The values derived by Feldt, that are reproduced in the table below were chosen so as to minimize the apparent imprecision. The table shows the optimal number of levels of a single blocking variable in terms of the total size of the experiment. N is the total number of subjects in the experiment and we assume that each cell of the design has an equal number of observations, or as nearly equal as possible for the selected values of N . The table also shows the correlation between the blocking and criterion variable (r), and the number of levels of the treatment dimension (i.e., either two or five as examples).

Table: *Optimal numbers of levels for a blocking variable*

| r | levels of A | n_t | | | | | |
|-----|-------------|-------|----|----|----|-----|-----|
| | | 20 | 30 | 50 | 70 | 100 | 150 |
| .2 | 2 | 2 | 3 | 4 | 5 | 7 | 9 |
| | 5 | 1 | 2 | 2 | 3 | 4 | 6 |
| .4 | 2 | 3 | 4 | 6 | 9 | 13 | 17 |
| | 5 | 2 | 3 | 4 | 5 | 7 | 10 |
| .6 | 2 | 4 | 6 | 9 | 13 | 17 | 25 |
| | 5 | 2 | 3 | 5 | 7 | 9 | 14 |
| .8 | 2 | 5 | 7 | 12 | 17 | 23 | 25 |
| | 5 | 2 | 3 | 5 | 7 | 10 | 15 |

Three features in the above table are readily apparent. First, for experiments having a larger number of subjects, there is an increase in the optimal number of levels for the blocking variable. Second, as the correlation between the blocking and criterion variables increases, the optimal number of blocks increases. And third, greater numbers of levels for blocks are optimal for smaller numbers of levels of the treatment dimension.

A randomized block applied to an independent variable with two levels (i.e. two groups) is called the paired-difference experiment. It is the simplest of the "noise-reducing" experiments and, in particular, it is a simple example of a randomized block design. Making comparisons within the relatively homogeneous conditions of a piece of material increased the quantity of information in the experiment. The only difference between that example and the general case is that the randomized block design may be applied to more than two treatments and, in theory, to as many as we please. In practice, as we have seen above, the number of treatments is kept to a minimum to maintain the within-block homogeneity in which a comparison of treatments is to be made. As the block size increases, the within-block variability tends to increase.

Suppose that we wish to compare the wearing quality of two types of automobile tires using a 20,000-mile road test as a basis for comparison. If only the rear wheels of each automobile are to be used and ten automobiles are to be employed in the experiment, two methods of design might be suggested. We could randomly assign ten tires each of type A and type B to the twenty wheels. This procedure, known as a *completely randomized design* has the obvious disadvantage that the wear measurements would vary greatly from automobile to automobile depending upon the driver, his method of accelerating and decelerating, and the road surface to which the tires were exposed. A better method of design would be to block out the driver-automobile noise by making comparisons within each automobile. Thus the automobile would be a block containing two experimental units (the two rear wheels). One, each of tire types A and B would be randomly assigned to the rear wheels of each automobile, yielding a randomized block design consisting of ten blocks and two treatments.

People may often be viewed as blocks. For example, suppose that we wish to compare the average length of time for four methods of assembling a device in a manufacturing plant. Should we assign a fixed number of people (say eight.) to each method of assembly and record the length of time each takes to assemble the device? This would follow a completely randomized design. Most likely, we would recognize the variation in the physical dexterity of individuals and attempt to block this noise out of the experiment. To do this we would have each person assemble the device

according to all four methods, A, B, C, and D, assigning the sequence of assembly in a random order. Thus each person would be a block, and comparisons of the assembly times for the four methods would be made within the relatively stable conditions provided by each person. As a final comment, we note that we are assuming that the assemblers do not become fatigued during the sequence of four assemblies and that their assembly time is not thereby increased as they move from the first to the last.

Pieces of test equipment often vary greatly in their characteristics. If a number of pieces of test equipment are used to test for differences in the average response of three different types of a device (A, B, and C), each piece of equipment should be assigned one each of A, B, and C and be tested in a random order. This procedure would block on pieces of test equipment, and eliminate their variability in making a comparison between the average yields of A, B, and C.

As a final example, we might note that a very common source of variability in an experiment is time. Various uncontrolled variables seem to create variations in the response of interest during the time duration of the experiment. Suppose that we plan to compare five different treatments, A, B, C, D, and E, and wish to apply each to four experimental units. If the experiments must be run sequentially in time, we would divide time into four blocks, testing one each of A, B, C, D, and E in each of the four time blocks in a random order. Thus, we might be able to run only five tests per day. The day (or block) would be subdivided into five time periods which would randomly be assigned to A, B, C, D, and E. The entire experiment would require four blocks (days). Comparisons of the treatments would be made within each day, thereby eliminating any day-to-day variability.

Summary

A researcher can use the randomized block design when the following conditions are met:

1. There is at least one independent variable with two or more treatment levels.
2. The variability of the experimental unit(s) within a block is more homogeneous (less variability) than among blocks
3. There is the same number of experimental units within each block
4. Treatment conditions are randomly assigned to the experimental units within each block.

Homogeneity within blocks can be done through matching, repeated measurements on the same experimental unit or using litters.

The linear model for the randomized block in a fixed-effects model would look like:

$$Y_{ij} = \mu + \alpha_j + \pi_i + \varepsilon_{ij}$$

where μ = grand mean (constant for all observations)

α_j = treatment effect (constant for all observations within treatment condition j).

$$\sum_{j=1}^k \alpha_j = 0$$

π_i = the blocking effect (constant within each block, i) $\sum_{i=1}^n \pi_i = 0$

ϵ_{ij} = experimental error. All ϵ_{ij} 's are independent of each other and normally distributed with mean = 0 and variance = σ_e^2

Since there is only one experimental unit per cell of the design, the interaction term between treatments and blocks, i.e., $\alpha\pi$, are additive. That is, there is no interaction effect. The absence of interaction between treatment and blocks implies that the covariances between all pairs of treatments (i.e. A-B, A-C, B-C, etc.) are equal. This is a very restrictive requirement. According to Kirk, the use of matched participants within a block would be more likely to satisfy this condition than the use of the same participant in a block.

For the fixed effects model, the expected mean square s would be:

$$\frac{E(MS_{treatment})}{E(MS_{residual})} = \frac{\sigma_e^2 + n\sigma_\alpha^2}{\sigma_e^2 + \sigma_{\alpha\pi}^2} \qquad \frac{E(MS_{block})}{E(MS_{residual})} = \frac{\sigma_e^2 + k\sigma_\pi^2}{\sigma_e^2 + \sigma_{\alpha\pi}^2}$$

If the assumption about equal covariances is true, the term: $\sigma_{\alpha\pi}^2$ will be equal to zero.

The additivity requirement in the fixed effects model is not so important in the random or mixed effects models.

The ANOVA Summary Table looks like the following:

| Source | Sums of Squares | Df | Mean Square | F |
|--------------------|---------------------|----------------|--------------------------------------|---------|
| Between Treatments | SSA | k - 1 | SSA/dfA | MSA/MSR |
| Between Blocks | SSB | n - 1 | SSB/dfB | MSB/MSR |
| Treat × Block | SS _{A×B} | (n - 1)(k - 1) | SS _{A×B} /df _{A×B} | |
| Total | SS _{Total} | N - 1 | | |

The computational formulas are

$$SS_{total} = SS_{treatment} + SS_{blocks} + SS_{A \times B}$$

$$SS_{total} = \sum_{i=1}^n \sum_{j=1}^k (X_{ij} - GM)^2 = \sum_{i=1}^n \sum_{j=1}^k X_{ij}^2 - \left(\frac{\sum_{i=1}^n \sum_{j=1}^k X_{ij}}{N} \right)^2 = SD^2(N)$$

$$SS_{\text{blocks}} = \frac{\sum_{i=1}^n B_i^2}{k} - \left(\frac{\sum_{i=1}^n \sum_{j=1}^k X_{ij}}{N} \right)^2 = \frac{\sum_{i=1}^n B_i^2}{k} - M^2(N)$$

$$SS_{\text{treatment}} = \frac{\sum_{j=1}^k A_j^2}{n} - \left(\frac{\sum_{i=1}^n \sum_{j=1}^k X_{ij}}{N} \right)^2 = \frac{\sum_{j=1}^k A_j^2}{n} - M^2(N)$$

The degrees of freedom will be $n - 1$ and $(N - n - k + 1)$.

Example: Eight judges completed a questionnaire designed to evaluate classroom teaching. Four films were shown to the judges. Each film portrayed different teaching performances. For each judge, the viewing of the four films was presented in random order. The data are given below:

| Films | Judges | | | | | | | | SUM |
|-------|--------|----|----|----|----|----|----|----|-----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| A | 9 | 10 | 7 | 5 | 12 | 7 | 8 | 6 | 64 |
| B | 4 | 9 | 3 | 0 | 6 | 8 | 2 | 4 | 36 |
| C | 12 | 16 | 10 | 9 | 11 | 10 | 10 | 14 | 92 |
| D | 9 | 11 | 7 | 8 | 12 | 7 | 7 | 8 | 69 |
| SUM | 34 | 46 | 27 | 22 | 41 | 32 | 27 | 32 | 261 |

For Treatment (Film):

$$H_0: \mu_A = \mu_B = \mu_C = \mu_D$$

$$H_1: \mu_i \neq \mu_j, \text{ for some } i, j \text{ where } i = 1, 2, 3; j = 2, 3, 4.$$

For Blocks (Judges):

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6 = \mu_7 = \mu_8$$

$$H_1: \mu_i \neq \mu_j, \text{ for some } i, j \text{ where } i = 1, 2, \dots, 7 \text{ and } j = 2, 3, \dots, 8.$$

$$M = 261 \div (8 \times 4) = 8.15625$$

$$M^2(N) = (8.15625)^2 (32) = 66.5244 (32) = 2128.78125$$

$$SD = 3.3737$$

$$SD^2(N) = 364.21875$$

$$SS_{\text{Total}} = 364.219$$

$$SS_{\text{blocks}} = \frac{\sum_{i=1}^n B_i^2}{k} - M^2(N) = \frac{34^2 + 46^2 + 27^2 + 22^2 + 41^2 + 32^2 + 27^2 + 32^2}{4} - 2128.78125$$

$$= (8943/4) - 2128.78125 = 2235.75 - 2128.78125 = 106.9688$$

$$SS_{\text{treatment}} = \frac{\sum_{j=1}^k A_j^2}{n} - M^2(N) = \frac{64^2 + 36^2 + 92^2 + 69^2}{8} - 2128.78125$$

$$= 2327.125 - 2128.78125 = 198.34375$$

$$SS_{A \times B} = 364.219 - 106.9688 - 198.34375 = 58.9065$$

ANOVA SUMMARY TABLE

| Source | Sums of Squares | df | Mean Square | F |
|---------------------|-----------------|----|-------------|-------|
| Between Films | 198.344 | 3 | 66.115 | 23.57 |
| Between Judges | 106.969 | 7 | 15.281 | 5.448 |
| Film \times Judge | 58.907 | 21 | 2.805 | |
| Total | 364.219 | 31 | | |

Critical value for Treatment (Films) $v_1 = 3$, $v_2 = 21$, $\alpha = .05$, $F_{\text{crit}} = 3.07$.

Critical value for Blocks (Judges) $v_1 = 7$, $v_2 = 21$, $\alpha = .05$, $F_{\text{crit}} = 2.49$.

Since $F_{\text{Treat}} = 23.57 > F_{\text{crit}} = 3.07$, reject H_0 .

Since $F_{\text{Block}} = 5.448 > F_{\text{crit}} = 2.49$, reject H_0 .