

4.2: Randomized Block Design

In randomized block design, the control technique is done through the design itself. First the researchers need to identify a potential control variable that most likely has an effect on the dependent variable. Researchers will group participants who are similar on this control variable together into blocks. This control variable is called a blocking variable in the randomized block design. The purpose of the randomized block design is to form groups that are homogeneous on the blocking variable, and thus can be compared with each other based on the independent variable.

How to Carry Out a Randomized Block Design

Using the example from the last section, we are conducting an experiment on the effect of cell phone use (yes vs. no) on driving ability. The independent variable is cell phone use and the dependent variable is driving ability. A potential control variable would be driving experience as it most likely has an effect on driving ability. Driving experience in this case can be used as a blocking variable. We will then divide up the participants into multiple groups or blocks, so that those in each block share similar driving experiences. For example, let's say we decide to place them into three blocks based on driving experience - seasoned; intermediate; inexperienced. You may wonder how we decide on three blocks. We will get to this in a little bit.

Once the participants are placed into blocks based on the blocking variable, we would carry out the experiment to examine the effect of cell phone use (yes vs. no) on driving ability. Those in each block will be randomly assigned into either treatment conditions of the independent variable, cell phone use (yes vs. no). As we carry out the study, participants' driving ability will be assessed. We can determine whether cell phone use has an effect on driving ability after controlling for driving experience. Again, since seasoned drivers are randomly assigned into both cell phone use conditions, as well as those with intermediate driving experience, and little driving experience, we already took care of the effect of the blocking variable, driving experience; so we are confident varied driving experience is not competing with the independent variable, cell phone use, in explaining the outcome variable, driving ability.

No Blocking Variable vs. Having a Blocking Variable

Randomized block design still uses ANOVA analysis, called randomized block ANOVA. When participants are placed into a block, we anticipate them to be homogeneous on the control variable, or the blocking variable. In other words, there should be less variability within each block on the control variable, compared to the variability in the entire sample if there were no control variable. Again going back to the same example, seasoned drivers may still vary in their driving experiences, but they are more similar to each other, thus as a subgroup would have less variability in driving experience than that of the entire sample. This is the key advantage of randomized block design. Less within-block variability reduces the error term and makes estimate of the treatment effect more robust or efficient, compared to without the blocking variable.

Without the blocking variable, ANOVA has two parts of variance, SS intervention and SS error. All variance that can't be explained by the independent variable is considered error. By adding the blocking variable, we partition out some of the error variance and attribute it to the blocking variable. As a result, there will be three parts of the variance in randomized block ANOVA, SS intervention, SS block, and SS error, and together they make up SS total. In doing so, the error variance will be reduced since part of the error variance is now explained by the blocking variable. In F tests, we look at the ratio of effect and error. When the numerator (i.e., error) decreases, the calculated F is going to be larger. We will achieve a smaller P obtained value, and are more likely to reject the null hypothesis. In other words, good blocking variables decrease error, which increases statistical power.

While it is true randomized block design could be more powerful than single-factor between-subjects randomized design, this comes with an important condition. That is we must select good blocking variables. As you have seen from the procedure described above, it shouldn't come as a surprise that it is very difficult to include many blocking variables. For one, the procedure becomes cumbersome. Also, as the number of blocking variables increases, we need to create more blocks. Each block has to have a sufficient group size for statistical analysis, therefore, the sample size can increase rather quickly. The selection of blocking variables should be based on previous literature.

Furthermore, as mentioned early, researchers have to decide how many blocks should there be, once you have selected the blocking variable. We want to carefully consider whether the blocks are homogeneous. In the case of driving experience as a blocking variable, are three groups sufficient? Can we reasonably believe that seasoned drivers are more similar to each other than they are to those with intermediate or little driving experience? It is a subjective decision left up to the researchers. If the blocks aren't homogeneous, their variability will not be less than that of the entire sample. In that situation, randomized block design can

decreases the statistical power and thus be worse than a simple single-factor between-subjects randomized design. Again, your best bet on finding an optimal number of blocks is from theoretical and/or empirical evidences.

Assumptions of Randomized Block Design/ANOVA

Randomized block ANOVA shares all assumptions of regular ANOVA. There are two additional assumptions unique to randomized block ANOVA.

First, the blocking variable should have an effect on the dependent variable. Just like in the example above, driving experience has an impact on driving ability. This is why we picked this particular variable as the blocking variable in the first place. Even though we are not interested in the blocking variable, we know based on the theoretical and/or empirical evidence that the blocking variable has an impact on the dependent variable. By adding it into the model, we reduce its likelihood to confound the effect of the treatment (independent variable) on the dependent variable. If the blocking variable (or the groupings of the block) has little effect on the dependent variable, the results will be biased and inaccurate. We are less likely to detect an effect of the treatment on the outcome variable if there is one.

Second, the blocking variable cannot interact with the independent variable. In the example above, the cell phone use treatment (yes vs. no) cannot interact with driving experience. This means the effect of cell phone use treatment (yes vs. no) on the dependent variable, driving ability, should not be influenced by the level of driving experience (seasoned, intermediate, inexperienced). In other words, the impact of cell phone use treatment (yes vs. no) on the dependent variable should be similar regardless of the level of driving experience. If this assumption is violated, randomized block ANOVA should not be performed. One possible alternative is to treat it like a factorial ANOVA where the independent variables are allowed to interact with each other.

This page titled [4.2: Randomized Block Design](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by [Yang Lydia Yang](#).