

LA-UR-17-31288 (Accepted Manuscript)

QP Statistics Spotlight: Randomization – The Rationale behind randomization, and options when up against constraints

Anderson-Cook, Christine Michaela

Provided by the author(s) and the Los Alamos National Laboratory (2019-04-03).

To be published in: Quality Progress

DOI to publisher's version:

Permalink to record: <http://permalink.lanl.gov/object/view?what=info:lanl-repo/lareport/LA-UR-17-31288>

Disclaimer:

Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by Triad National Security, LLC for the National Nuclear Security Administration of U.S. Department of Energy under contract 89233218CNA000001. By approving this article, the publisher recognizes that the U.S. Government retains nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or to allow others to do so, for U.S. Government purposes. Los Alamos National Laboratory requests that the publisher identify this article as work performed under the auspices of the U.S. Department of Energy. Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.

At Random

The rationale behind randomization, and options when up against constraints

by Christine M. Anderson-Cook

When I was a faculty member in the department of statistics at Virginia Tech more than a decade ago, I sometimes taught the “Statistical Methods for Scientists and Engineers” course. This course was primarily taken by graduate students from other departments who wanted to learn enough statistics to be able to effectively design and analyze an experiment as part of their thesis or dissertation work.

As part of the course, I always emphasized the importance of randomization to protect against the unexpected and to ensure the validity of the interpretation of results for establishing causality.

After taking the stats methods class, one chemical engineering graduate student made an appointment with me in the statistics consulting center. He wanted help designing an experiment for his master’s thesis. After discussing the experiment’s goals, which factors to manipulate and their ranges, I provided him with a designed experiment (in this case, a 16-run 2^4 factorial design) that matched his needs.

As was my usual habit, we talked through the design characteristics with the design in standard order (Figure 1—part a), and I provided another version of the design in randomized order (Figure 1—part b). I began to give him instructions about the implementation of the proposed design based on the randomized design table. He interrupted and said: “Oh, come on. You don’t really expect me to do this experiment with that randomization thing. That’s just some theory that you statisticians tell us in classes.”

I must have looked suitably shocked: This was a student who had done well in my class, and by all accounts had taken in my admonition to randomize experiments seriously when completing class assignments and tests. I asked, “Why would randomization not be relevant to

your experiment? Don't you remember the advantages of randomization that we discussed in class?"

In his mind, the "untidy run order" from randomization seemed unorganized, unscientific and something that a good researcher would never do. So, he and I spent the next half hour talking about the merits of randomization, and how it was absolutely relevant and necessary for good scientific investigation.

Merits of randomization

I'd like to recap some of these principles, as well as provide some options when faced with constraints on randomization when implementing an experiment. The design shown in Figure 1 is called a *completely randomized design* because there were no restrictions in determining the order of the experimental runs.

The same 16 runs are shown in Figure 1 (parts a and b), but the run order for the latter has been permuted from the standard order—shown in Figure 1 (part a). When it is run, the assumption is that between each run, the levels of each factor are reset. For example, for the first run, factors X1 and X2 at their low levels, and X3 and X4 at their high levels. For the second run, X1 is set at its high level and X2 at its low level, and so on. For factor X1, this involves changing the setting from low to high, while for X2, the setting stays the same level, but should still be set again. Regardless, the assumption is that the levels for each of the five factors will be reset between each run.

Why is this beneficial? Suppose that factor X2 is a temperature with levels 200° F (for the low level denoted with -1) and 250° F (high level denoted +1), and for the first run, the temperature was actually set at 199° F. By resetting the temperature between each run, the potential for systematic bias can be reduced. If the factors were not reset after each run, then runs one, two (and three and four, which all of have X2 at the low setting) could be run at the slightly low temperature. In addition, it gives us information about how much variability is associated with the setting of each factor, which might have an impact on the response values obtained.

There are other reasons that randomization of the order of experimental runs is important: The first is related to the fact that there are other things changing during the course of running the experiment. Perhaps unbeknownst to the experimenter, the humidity in the lab is slowly increasing while the experiment takes place, and suppose that humidity affects the response.

If the experiment was run in the order shown in Figure 1 (part a—standard order), the first eight runs would be performed at the low level of X1, and the last eight runs would be at the high level of X1. If humidity is rising and is affecting the response, it will look like factor X1 is the explanation for the changes to the response that are actually being driven by humidity. This can lead to us falsely concluding that changes in X1 cause changes in the response.

If the randomized run order in Figure 1 (part b) is used, however, notice how none of the factors have too many runs at either the high or low level in either the first or the last half of the experiment. This offers protection from falsely associating a change in response driven by a lurking variable (one that is not being intentionally manipulated by the experimenter) with a factor in the experiment.

Would it be better to have been tracking humidity or, better yet, holding it constant throughout the experiment? Yes, but that assumes that we can anticipate everything that will affect the response and that we have the ability to control all of these aspects. With randomization, however, we might not realize that humidity is important, but at least it won't contaminate our interpretation of the relationship between the studied factors and our response.

Finally, statistical theory for the analysis depends on randomization. The interpretation of p-values from an analysis of variance or other analyses of the designed experiment is tied to the assumption of the run order and experimental units being randomly assigned.¹ This theoretical underpinning means that we have the ability to assess and quantify how likely the change that we have observed for a particular factor or term in our model can be attributed to chance. The smaller that chance, the more confident that we feel about associating that observed change with the factor that we have manipulated.

When thinking about randomization in the experiment, it is also important to make sure that we:

1. Design the experiment that we want with the factor and factor levels that we want.
2. Fix whatever factors we are not manipulating.
3. Randomize over the run order of the experiment and the assignment of experimental units to the treatments.²

Always random?

So, does that mean that we should always run a completely randomized experiment? No—there are other types of designs to accommodate constraints and situations during experimentation, while still preserving the benefits of randomization. The following are three different alternatives and the situations for which they are suitable:

1. **Blocking designs:** These designs take into consideration that there are factors that cannot be randomly assigned, or factors that cannot be controlled during the experiment. If we run an experiment to determine the impact of different doses on the response of interest for our test subjects, for example, we cannot randomly assign gender. Hence, a good design would treat gender as a blocking factor, and assign factor combinations within the blocks.³

Alternatively, we might have an experiment (such as the earlier humidity example) in which conditions during the test vary over time. In this case, we could run a blocking design that considers different portions of the day as blocks. This imposes balance of treatment combinations in the blocks and also allows qualitative exploration of potential differences between the blocks.

Figure 2 shows a sample of how a 2^4 factorial experiment can be run in two blocks of size eight. The blocks might coincide with gender or time of day. Note how in each block, there are four runs at the high level and four runs at the low level of each factor. This ensures that factor effects can be estimated independently of the block effect. Hence, for these types of designs, the primary goal is to take into account a factor or conditions that cannot be controlled during the experiment, but are thought to potentially affect the response. The analysis of the experiment should include a term for the blocking because differences between the blocks should be removed from the estimate of the natural variability.⁴

2. **Split-plot designs:** These designs recognize the fact that some factors may be time-consuming, complicated or expensive to adjust, and allows for groups of runs to be performed without resetting all of the factor levels between the runs.^{5,6} The requirement that each factor be reset for each run of the experiment is often impractical, so split-plot designs allow for two (or more) levels of randomization.

Figure 3 (part a) shows a 16-run experiment in which factor X1 is hard to change, and hence the experimenter wishes to reduce the number of times that factor is reset. The experiment has four whole plots, where within each of these, the level of factor X1 is set just once and then four subplot runs are performed. For each of the four runs in each whole plot, the levels of the

easy-to-change factors X2, X3 and X4 are reset each time. In setting up the experiment, two randomizations are performed:

1. The order that the whole plots will be performed.
2. The order that the subplots are run within each whole plot.

Figure 3 (part b) shows another variation of the same design, but with a different randomized order. Note how the order of the whole plots has changed, as well as the order of the runs in each block. Split-plot designs can flexibly accommodate more than one hard-to-change factor, but the analysis of the experiment must reflect the two different randomizations, which each may have different associated amounts of natural variability.⁷ Deciding whether this is the right type of design for your experiment balances ease of implementation, potential reduced cost for fewer changes of the hard-to-change factors, and additional complexity of analysis.⁸

3. Sequential experimentation: Finally, another option to a completely randomized design is to run the experiment with the goal of getting partial information about the response as the experiment is being run. For example, consider a definitive screening design (DSD)^{9,10} with three three-level continuous factors and three two-level categorical factors. These DSDs are supersaturated designs, but are capable of separately estimating all main effects, two-factor interactions and quadratic terms.

In many experiments, it is beneficial to obtain preliminary information about the range of responses, as well as early feedback about the choice of factor levels and the factor effects on the response. By taking advantage of the notion of blocking, we can divide the experiment into two blocks, with the goal of building in a slight pause in the experiment after the first block has been completed.

At this time, we can take stock of how the experiment is running: Are we getting sensible results at each of the factor levels? Do the main effects for each factor match our expectations? Is the range of responses what is required for optimization or other post-experiment decisions?

If some of these answers do not match what was anticipated, the opportunity exists to make some adjustments to the experiment to avoid wasting further resources and maximizing the information obtained. If things are turning out as expected, the second block of the experiment can be completed, and only a small modification to the analysis (the inclusion of a blocking term in the model) is required.

Randomization is performed, as with any blocking experiment: The run order of experimental units in a block is randomized. This simple strategy of creating a small number of blocks—and strategically choosing to run one or more of them first—allows for easy design construction, while still being able to get preliminary answers to fundamental questions early in the experiment.

A note on choosing which blocks to run first: In the DSD case, only block one has all of the factors at all of their levels. So if this is one of the objectives to assess early in the experiment, choosing block one to implement first is important. If several blocks satisfy this objective, their relative D-efficiency for the main effects model can be considered when choosing between alternatives. Sequential experimentation is an important aspect to consider when designing an experiment: If we are able to learn how the experiment is going early on, we can adapt if needed to maximize the information that we are able to gain.

While randomization may seem like a small part of the overall plan for implementing a designed experiment, it is important to protect against systematic bias and to justify the interpretation of the analysis results.

A completely randomized experiment is the simplest way to include this in your experiment plan, but blocking or split-plot designs provide alternatives to match the needs of the study.

REFERENCES

1. Klaus Hinkelmann and Oscar Kempthorne, *Design and Analysis of Experiments, Vol. 1: Introduction to Experimental Design*, Wiley-Interscience, 1994, pp. 162-168.
2. Christine M. Anderson-Cook, “What and When to Randomize,” *Quality Progress*, March 2006, pp. 59-62.
3. Douglas C. Montgomery, *Design and Analysis of Experiments*, eighth edition, Wiley, 2012, pp.139-157.
4. Raymond H. Myers, Douglas C. Montgomery and Christine M. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley-Interscience, 2016, p. 139 and p. 142.

5. Bradley Jones and Christopher J. Nachtsheim, "Split-Plot Designs: What, Why and How," *Journal of Quality Technology*, Vol. 41, No. 4, 2009, pp. 340-361.
6. Peter Goos, *The Optimal Design of Blocked and Split-Plot Experiments*, Springer, 2012.
7. Myers, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, see reference 4, pp. 141-145.
8. Christine M. Anderson-Cook, "When Should You Consider a Split-Plot Design?" *Quality Progress*, October 2007, pp. 57-59.
9. Bradley Jones and Christopher J. Nachtsheim, "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects," *Journal of Quality Technology*, Vol. 43, No. 1, 2011, pp. 1-15.
10. Bradley Jones and Christopher J. Nachtsheim, "Definitive Screening Designs With Added Two-Level Categorical Factors," *Journal of Quality Technology*, Vol. 45, No. 2, 2013, pp. 121-129.

Christine M. Anderson-Cook is a research scientist in the Statistical Sciences Group at Los Alamos National Laboratory in Los Alamos, NM. She earned a doctorate in statistics from the University of Waterloo in Ontario, Canada. Anderson-Cook is a fellow of ASQ and the American Statistical Association. She is the 2018 recipient of the ASQ Shewhart Medal