Experiments, Randomization, and Observational Studies

Donald J. Wheeler

When analyzing data it is essential to distinguish between *observational studies* and *experimental studies*. The data may be said to have come from an observational study when they arise as a side effect of some continuing operation or on-going process. In contrast to this, when a series of operations are carried out specifically in order to obtain specific data, those data may be said to have come from an experimental study. Virtually all data can be said to come from either an observational study or an experimental study.

Clearly, process behavior charts were intended for use with observational studies. Since it will generally be much easier to obtain data as a side effect of operations, this alone makes the process behavior chart one of the most important tools in your data analysis tool kit. However, as is shown in the several of my books, process behavior charts may also be used with experimental data. In fact, they will often be the only analysis technique that will reveal certain aspects of your data (such as a lack of homogeneity where the data should be homogeneous). However, by the very nature of an experiment, experimental data will be collected under two or more conditions. This characteristic of experimental data places restrictions on how the data may be analyzed.

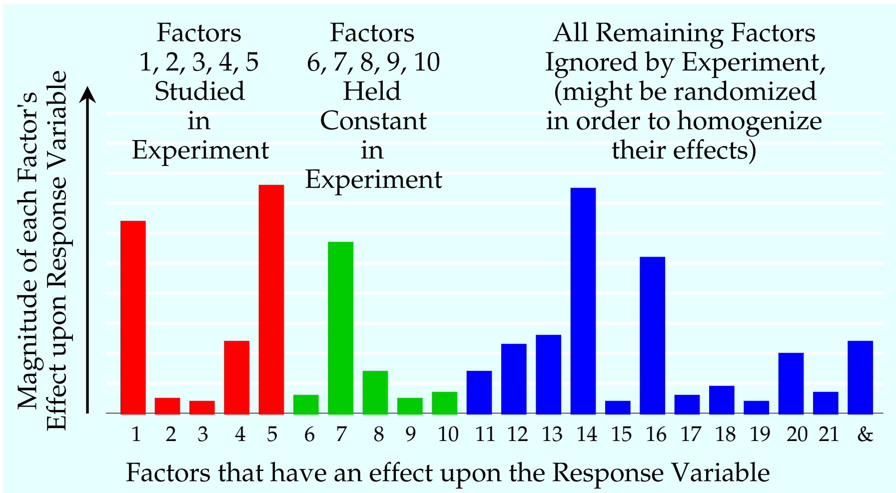
Experiments

Every experiment will have the following characteristics. A response or outcome variable will be defined. Out of the large collection of input variables that may have some impact upon this response, a subset, consisting of a few variables, will be selected to be studied in the experiment. The remainder of the input variables are then excluded from the study. If we think that some of these extraneous factors may have an impact upon the response, then we might hold these factors constant during the course of the experiment, however, most of these extraneous factors will be ignored during the course of the experiment. Thus there are three things we can do with an input variable during the course of an experiment, we can study it, we can hold it constant, or we can ignore it.

When we perform an experiment while holding one or more input variables constant at some level, our experimental results will only characterize what happens when the constant factors are at those particular levels. The experiment will not tell us what happens at other values of the constant factors. What if these fixed factors interact with the experimental factors? What if they interact with each other? Since the fixed factor levels do not allow you to detect these interactions, your experimental results will be of limited value. It is rather like trying to describe the life of a lion by watching the behavior of a lion in a small cage. While you may observe that lions sleep a lot, sometimes pace around, and can make some very loud sounds, you will not learn much about how lions behave in the wild by observing them in a cage. Your observations about lions may be correct, but they will not be complete.

But what about the factors that are simply ignored by the experiment? When we ignore a factor we are making an implicit assumption that that factor does not have a pronounced effect upon the response variable. But how can we ever be sure that this assumption is correct?

The honest answer is that we cannot be sure that the factors we have ignored or held constant do not have important effects upon the response variable.



What might you learn from this experiment?

What are you going to miss with this experiment?

Randomization

The problem of ignoring factors that had appreciable effects upon the response variable haunted all attempts to use the scientific method with agricultural and biomedical problems at the end of the Nineteenth Century. In agricultural and biometric studies the basic experimental units are either plots of land or animals. Given that all such experimental units are different, and are subject to a myriad of influential factors that are beyond the control of the experimenter, how could anyone ever make sense of a set of experimental data? The answer that emerged was to randomly assign the treatment combinations (experimental conditions) to the various experimental units (fields, animals, etc.) and then to average out the effects of the extraneous influential factors (within each treatment). Differences that showed up between the various treatment averages could then be said to be due to the treatments themselves.

Variations of this approach soon became standard in agricultural and biometric research. Randomization provided an insurance policy that one treatment would not get all of the “best,” or all of the “worst,” experimental units, and thereby be erroneously judged to be better or worse than the other treatments. Randomization protects against various forms of bias that can creep into an experiment, even indirectly, and therefore it has become a very useful research tool.

However, it is important to note how and when randomization works. The mechanism used by randomization is that of averaging. When an experiment uses each treatment combination several times, the randomization of which experimental units receive each treatment combination, or the randomization of the order in which the different treatment combinations are studied, will tend to average out the effects of the various factors that are not included in (and are therefore ignored by) the study. (At least that is the way it should work most of the time. In any one experiment there is no guarantee that randomization will prevent the effect of an ignored factor from showing up in your analysis.) Since the basic mechanism is averaging, it should be apparent that randomization becomes less effective as the number of observations per treatment combination gets smaller. This will happen simply because, as the subgroup size gets smaller, there will be less chance for the extraneous effects to average out within each subgroup.

Secondly, when does randomization work? Randomization works when you cannot resort to the primary confirmatory tool of the scientific method—the nontrivial replication of results. Virtually all scientific and engineering knowledge has been gained by virtue of the replication of results. While randomization is central in agricultural and biomedical research, it is virtually unheard of in physics, chemistry, and engineering. The reason for this is seen in the nature of experiments in the different areas. In the hard sciences experiments tend to occur sequentially in time. In agricultural and biometric studies it will often take a whole season, or even a whole generation, to conduct an experiment. Since waiting for several seasons, or several generations, to conduct a series of experiments in order to verify a finding is not realistic, experiments in these areas tend to build the replication into the study itself by using multiple experimental units all at the same time. When this is done, randomization is what makes the replication nontrivial. Thus, the experiment and its replication are done simultaneously.

In other words, when you are performing an experiment where each treatment will be applied to a large number of experimental units, and when you will not have the opportunity to conduct subsequent, confirmatory experiments, then randomization is both an insurance policy and a necessary part of good scientific experimentation. When we are doing basic research, where we are not concerned with the identification of *all* the relationships that exist, but merely with the existence of certain selected relationships, this approach makes sense. It is sound and it is proven.

Observational Studies

However, if you are concerned with optimizing a system, or getting a process to work, you cannot depend solely upon experimental results which are always obtained in a limited context. You have to deal with the response variable in the presence of all of the factors that have an impact. You cannot simply study some factors and ignore the others. But, of necessity, every experiment will choose some factors and exclude other factors. So while you may begin with a set of experiments, you need to remember that limited results and conditional relationships do not tell the whole story. Eventually you will need a holistic approach, and this is what observational studies provide. In contrast to experiments, observational studies are like watching lions in their natural habitat rather than studying them in a cage. It may take longer to discover things with an observational study, but all the possible interactions and all the various factors are present and are allowed to make their contribution to the results of the study. When a factor makes its presence known in an observational study, you can be sure that it is a dominant factor.

With an observational study the clues to the source of any particular behavior will come from the context for each observed event. Here the key to discovery is the connection between context and the observed behavior. Data have to be analyzed and interpreted in terms of their context. Moreover, since we are not ignoring any of the input variables, there is no need for any insurance device, like randomization. In fact, any attempt to impose randomization on an observational study will merely result in confusion. In an observational study some of the most important information may consist of the time order sequence for the data. Therefore, with an observational study, any careful data analysis must preserve the time-order sequence of the data.

While we may learn some things by observing a lion in a cage, our knowledge can be made more complete by observing lions in the wild.

This article originally appeared in ***SPC Ink, February 2005***

Dr. Wheeler is a Fellow of both the American Statistical Association and the American Society for Quality. As the author of dozens books and hundreds of articles, he is one of the leading authorities on SPC and Applied Data Analysis. Contact him at www.spcpress.com.