Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Statistical Planning and Inference 133 (2004) 199–210

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Randomized and random run order experiments

Jitendra Ganju[a,*], James M. Lucas[b]

[a] Chiron Corp., 4560 Horton Street, M/S U-140, Emeryville, CA 94608, USA
[b] J.M. Lucas and Associates, 5120 New Kent Rd., Wilmington, DE 19808, USA

Received 23 January 2003; accepted 15 March 2004
Available online 9 June 2004

Abstract

Randomization in industrial and scientific experiments on equipment has meant randomizing the order of application of levels of treatments to units. This definition is inadequate because it does not render independent error terms. Randomization also requires independent resettings of treatment levels when the levels for the preceding run are the same. We review how the literature incorrectly explains how randomization is to be carried out. The need to reset levels of a treatment from one run to the next is never emphasized. Using a simple example we show why statistical tests are biased for all treatments even when levels for just one treatment are not independently reset. Even if the expected mean squares recognize the restrictions on randomization, the usual F test will not give predictable results because its numerator and denominator are correlated.

Experimental design on equipment includes experiments from the chemical, automobile, pharmaceutical, and aeronautical industries. The statistical interpretation of data from such experiments will be misleading. Books on experimental design must emphasize the independent resetting of levels just as carefully as they emphasize the random assignment of treatment levels.

MSC: 62K05; 62K10; 62K15; 62K20; 62K25

Keywords: Inadvertent split-plotting; Industrial experimentation; Resetting of levels

1. Introduction

Remarkably, despite the widespread use of randomization, its meaning in almost all experimental design books is either inadequate or incorrect for scientific and industrial experiments on equipment because they do not discuss the need to reset treatment or

* Corresponding author. Tel.: +1-510-923-6418; fax: +1-510-923-3450
E-mail address: jitendra_ganju@chiron.com (J. Ganju)

0378-3758/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2004.03.009

factor levels independently from run to run. Many scientific and industrial experiments are performed using a random run order (RRO) yet when successive runs of a factor have the same level that factor is not reset. Therefore (complete) randomization is not achieved and the errors will not be independent. In fact, hypothesis tests will be biased and the numerator and denominator of F tests will usually be correlated. The simple fact is: if you do not reset (every factor level on each experimental run) you are not randomizing (and you cannot be running a completely randomized design). We, along with some co-workers, have been studying the properties and implications of running experiments when factors are not reset when successive runs use the same factor levels. In this article we will discuss randomization examples from the literature and present some new results that should be included in courses on experimental design. We use a simple example to explain what happens when factor levels are not reset on each run.

We begin by discussing an experiment from the book by Mason et al. (1989, Table 10.3). They discuss the design of a half-fraction of a 2-level, 6-factor experiment. The 6 factors, each of two types or of 2 levels, are piston ring, engine oil, engine speed, intake temperature, air–fuel mixture, and oxygen content. Let A and B denote the 2 types of piston rings. The authors present a random schedule for the *order* of application of the 32 treatment combinations and state that for the purposes of a *completely randomized analysis* “There are 20 changes of the piston rings...” The random order of application of the 32 piston rings in their example is (runs 1 through 32): A B A BBB AA B A B A BB A BB AAAA BB AA B AA BB A B. The 20 changes correspond to changing the type of piston rings only when the type of piston ring is different from one run to the next. This means that the piston ring is not independently reassembled if the same type is required for consecutive runs (the first occurrence of not resetting is at run 5, the second at run 6, the third at run 8, and so on). When successive runs of the experiment requiring the same type of piston rings are not independently reset, then the errors are correlated. This correlation violates the “independent” property of measurements that are believed to be “independently, and identically distributed”. One way to describe this phenomenon is to state that there is inadvertent split-plotting when factors are not reset.

Ganju (1994) and Ganju and Lucas (1997) prove that over all run orders the factor that is not reset gives biased tests and the other reset factors gives nearly unbiased tests. This article shows that the near unbiasedness of the reset factors is not of much comfort because for *any* run order, tests on *all* factors can be biased. Experiments that only randomize the run order of application of treatment combinations without independently resetting levels or changing types from run to run are not completely randomized experiments, and we refer to them as random run order or RRO experiments. An RRO experiment frees the estimation of treatment effects from some biases; for example, it guards against trends and cycles. Independent resettings after every run is needed to give independence errors necessary for permutation tests to provide the correct distribution of test statistics for the testing of hypotheses. When properly randomized, permutation-based tests and normal theory-based tests give similar results (Kempthorne and Doerfler, 1969).

It is remarkable that the literature is replete with emphasis on randomizing the *order* of application while the need for *independent resetting* of factor levels is almost always absent. Had awareness of the need for resetting existed, then it would have spawned a series of publications addressing new methods of running experiments because of the physical