Common Statistical Mistakes in Entomology:

Blocking and Inference Space

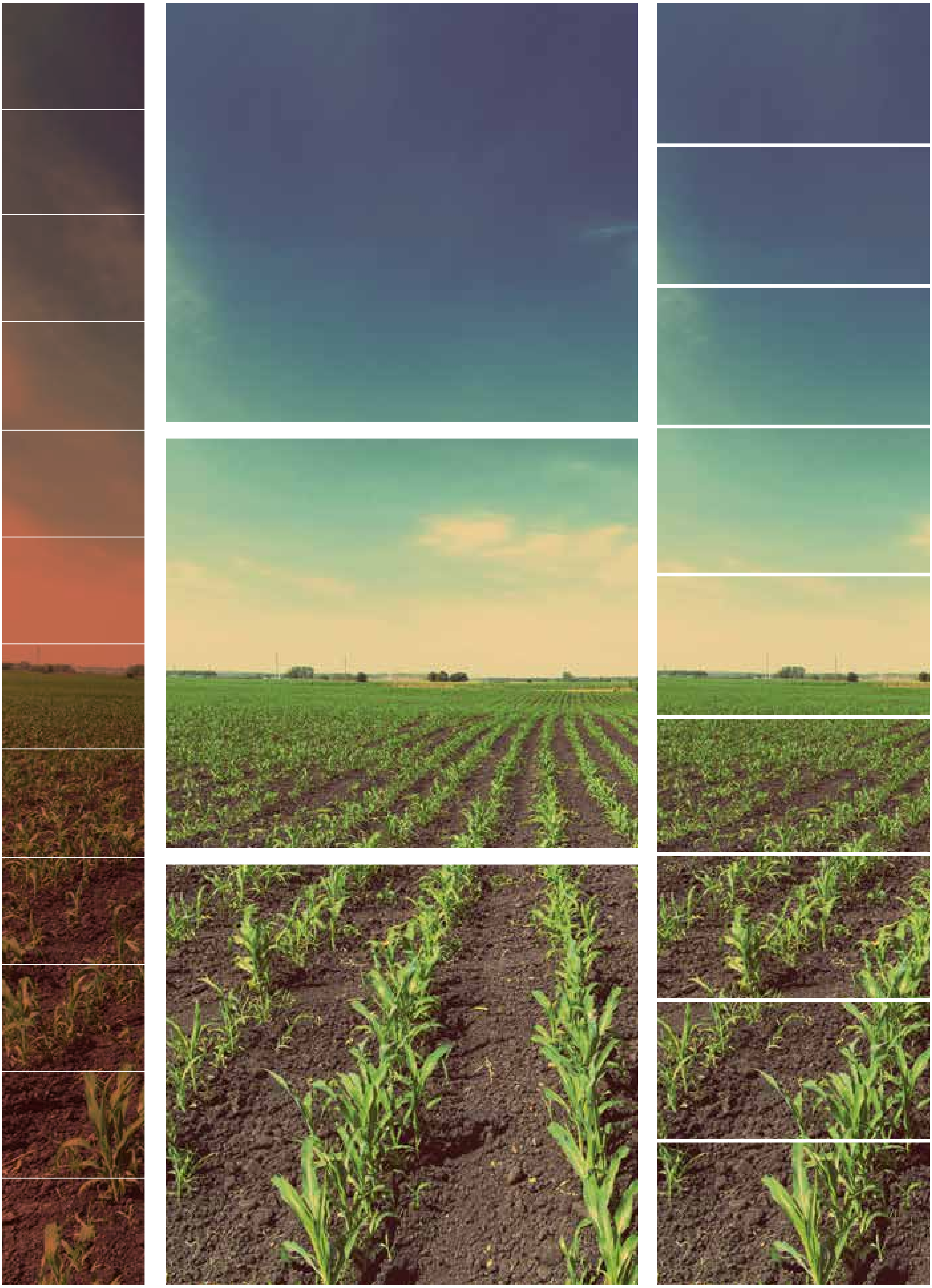
DALE W. SPURGEON

BLOCKING IS ... AN OPPORTUNITY TO DEFINE THE LARGER POPULATION OF CONDITIONS THAT DEFINES THE INFERENCE SPACE.

The final article in this series on common statistical errors focuses on the practice of blocking in experiments, and its implications on inference space; i.e., the range of situations to which the results are applicable. I address the concept of inference space first by contrasting fixed-effect and mixed-effect models. A fixed-effect model is one in which all of the model effects are assigned or controlled by the investigator, with specific interest in their impacts on the response. One chooses the factors investigated, and their levels, with that in mind. Importantly, there is no intent to draw inferences beyond the specific treatments and conditions that are observed. In a fixed-effects analysis, the statistical tests, variances, and standard errors are, strictly speaking, only applicable to the conditions and experimental units that were observed. In contrast, a mixed-effects model includes the fixed effects of interest, and one or more random effects that represent a sample from a larger population of physical entities or conditions. These entities (plots, blocks, fields, experimental repetitions) are not, in and of themselves, of interest; they simply represent a population of entities or conditions to which the results of the experiment are applicable. The inferences of a mixed-effects analysis are intended to extend to a larger population of experimental units and conditions similar to those that were observed. The test statistics and standard errors from a mixed-effects analysis reflect this broader inference space. Fixed-effect models are common in the current entomological literature. However, we usually intend our findings to be more broadly applicable, so the intended inference space is better aligned with a mixed-effects model. In practice, the breadth of the inference space depends to a great degree on the nature and arrangement of any blocking or repetition effects in the experiment.

Blocking, in a general sense, is intended to account for sources of variation that are influential but not of specific interest. When not addressed, these sources of variation reduce the precision with which one can estimate the variation associated with the treatments, or fixed effects. Blocks are clusters of experimental units that are alike, but that are presumed or known to be different

Editor's Note: This article is the fourth, and final, in a series of commentaries that address common statistical mistakes in entomology.



from the experimental units in other blocks. In that context, clustering relatively homogeneous experimental units into blocks allows us to account for variation in the response caused by differences among the blocks, thus improving our ability to detect differences caused by the treatments. A block is simply a non-treatment factor that partitions a meaningful source of variance, and may include sets of plots, experimental sites, cohorts of insects, or repetitions of an experiment.

Blocking is typically employed in field experiments as a source of replication; however, blocking is not always used effectively. Although it is important that the experimental units within a block are as alike as possible, it is not desirable for the blocks to be alike (Fig. 1). If the blocks are uniform, or nearly so, the benefit of blocking is lost. Sometimes blocking is used simply out of familiarity or convenience, and sometimes as a hedge against possible or anticipated inter-block gradients that have not yet appeared or that are difficult to identify. Repetition of an experiment is a good example of the latter. Blocking is also an opportunity to define the larger population of conditions that defines the inference space. In this context, it is important to recognize that as blocks become more alike, the inference space becomes narrower. Therefore, it is important that a sample of blocks represents the intended inference space.

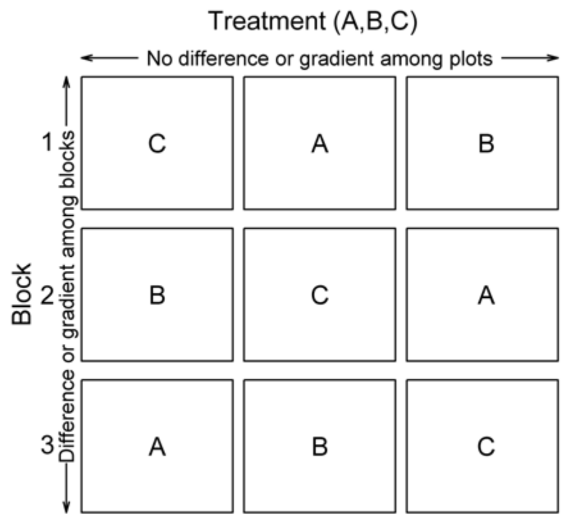


Fig. 1. A randomized block design illustrating the desired orientation of blocks to a source of variation (difference or gradient). Appropriate use of a blocked design is dependent on the existence of differences among blocks, relative homogeneity of plots within blocks, and absence of an interaction between blocks and the experimental treatments.

Implicit in any form of blocking is the lack of interaction between blocks and responses to the fixed effects or treatments. This assumption is consistent with the notion of dividing the experimental design into a treatment structure and a design structure (Milliken and Johnson 1984) that was discussed in a previous article (Spurgeon 2019a). The greater the differences among blocks, the broader the inference space. If the blocks are so different that treatment responses are not consistent and the treatment effect is not repeatable, the statistical inference of no treatment effect is correct, but only in reference to the inference space. In this situation, the inference space likely is too broad.

Consider, for example, an evaluation of trap designs for a beetle that exhibits a marked preference for one or more development stages of the host plant. Further assume that the locations of the beetles on the plants varies with crop-growth stage, which influences their response to the traps. The corn rootworm complex (*Diabrotica* spp., Coleoptera: Chrysomelidae) in corn (*Zea maize* L., Poales: Poaceae) would fit this description. In this scenario, responses to different trap designs will depend on the beetle population level, plant phenology, placement of the traps within the crop, and characteristics of the traps. If repetitions of the experiment are blocked within a field where plant development is uneven, or blocks are in fields of differing growth stages, the relative responses to the traps likely will exhibit an interaction between trap design and blocks. If the differences in host-plant development are too great, the inference space is too broad, because the responses to the trap designs are not consistent or repeatable over all of the growth stages. Although it might be tempting to evaluate the block and the treatment × block effects, unless treatments are replicated within blocks (a generalized randomized block design), there is no appropriate error term for F-tests of these effects. The solution is to expand the experimental design to encompass plant-growth stage as a fixed effect, or limit the blocks in time and space to plots or fields of the relevant growth stage(s). The point of this example is that the investigator must use knowledge of the subject area to devise a blocking scheme that is consistent with the desired inference space. If the intended inference space is not considered in the design process, the investigator leaves to chance that the resulting inference space will be too narrow for general use or too broad to detect differences.

A blocking effect often represents a composite of more than one factor. Consider, for example, sets of treatments established in different fields so that field is a blocking effect. The different fields may have different soil types, may be planted to different varieties, or may be managed differently. If these combinations of conditions are within the intended inference space, it is irrelevant that soil type, cultivar, and management scheme are confounded with the blocks. The point is that the blocks are different and the results reflect treatment responses within the range of those differences. Blocking can also be imposed on more than one level of the design structure. For example, a blocked design can be duplicated in different fields so that block and field are both random blocking effects. Blocking effects can be spatial, temporal, or even nested. What is important is that the blocking effects are selected so that they do not interact with the treatments, the blocks result in a meaningful inference space, and appropriate interaction effects are included in the analytical model to serve as error terms for tests of the treatments.

Blocking over time (repetition of experiments) can be a useful means of assessing the repeatability of results, as well as for obtaining treatment means and standard errors that reflect those that would be expected if the experiment was independently reproduced. This design requires that the experimental units observed in different repetitions are unique (independent). It is not uncommon that published reports focus on designs in which the same experimental units are observed multiple times, and time of observation (sample date) is considered a random, blocking effect. This approach is inappropriate because the experimental units within each block in time are not independent and cannot be properly considered as replicates of the treatments. This is because serial observations on the same experimental unit are often correlated, and appropriate analysis of these data requires that this correlation be modeled (a repeated-measures analysis). In field experiments, the time of observation may be important in and of itself, and may interact with the treatments. Treating observation time as a random blocking effect ignores this potential interaction and can lead to misinterpretation of the results, as described in a previous commentary (Spurgeon 2019b).

Blocking, whether physical or temporal, can improve the precision and power of experiments, so long as the blocking strategy is consistent with the intended inference space. If inference space was explicitly treated as a consideration during planning of the experiments, the occurrence of fixed-effect models in research reports would be considerably reduced. However, the concepts of blocking and inference space do not appear to be widely understood. In fact, a manuscript by the author of this commentary was recently rejected because the experiment was blocked (repeated) over time. The reviewer argued that the same results would not be expected from different repetitions of the experiment, and, therefore, by repeating the experiment, the author had confounded the results. This opinion illustrates a profound lack of understanding of blocking and inference space.

This series of commentaries addresses some of the most basic statistical problems that commonly occur in the entomological literature. Many other articles address more advanced problems, such as the misuse of hypothesis tests based on p-values (Yoccoz 1991), the growing body of published research that is poorly reproducible (e.g., Halsey et al. 2015), and the inappropriate or unnecessary transformation of non-normal data (O'Hara and Kotze 2010, Stroup 2013). Additional issues, such as analysis of non-Gaussian distributions, accommodating heterogeneous variances, and modeling correlated observations (repeated measures) can be addressed by modern statistical software, provided the basic experimental design is sound. These topics are important. However, if an experiment is designed in such a way that its analysis cannot provide valid hypothesis tests, then the independence or distribution of the errors and the apparent significance of the effects are irrelevant. I hope that by addressing the most basic analytical errors, this series of commentaries will promote a higher level of statistical literacy among researchers, thereby improving the reliability and repeatability of published results.

Acknowledgements

Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

References Cited

Halsey, L.G., D. Curran-Everett, S.L. Vowler, and G.B. Drummond. 2015. The fickle *P* value generates irreproducible results. Nature Methods 12: 179–185.

Milliken, G.A., and D.E. Johnson. 1984. Analysis of messy data, volume 1: designed experiments. Van Nostrand Reinhold, NY.

O'Hara, R.B., and D. Johan Kotze. 2010. Do not log-transform count data. Methods in Ecology and Evolution 1: 118–122.

Spurgeon, D.W. 2019a. Common statistical mistakes in entomology: models inconsistent with the experimental design. American Entomologist 64: 87–89.

Spurgeon, D.W. 2019b. Common statistical mistakes in entomology: ignoring interactions. American Entomologist 64: 192–195.

Stroup, W.W. 2013. Generalized linear mixed models: modern concepts, methods and applications. CRC Press, Boca Raton, FL.

Yoccoz, N.G. 1991. Use, overuse, and misuse of significance tests in evolutionary biology and ecology. Bulletin of the Ecological Society of America 72: 106–111.

Dale W. Spurgeon, *USDA, ARS, Arid-Land Agricultural Research Center, 21881 N Cardon Lane, Maricopa, AZ. E-mail: dwspurg@gmail.com*

DOI: 10.1093/ae/tmz064

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION

American Entomologist (ISSN 1046-2821) is published quarterly (4 issues per year) by Oxford University Press, a nonprofit organization located at 2001 Evans Road, Cary, North Carolina 27513-2009. The Editor is Kevin Steffey, Entomological Society of America, 3 Park Place, Suite 307, Annapolis, MD 21401-3722. There are no known bondholders, mortgagees, or other security holders owning or holding 1 percent or more of total amount of bonds, mortgages, or other securities. The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes have not changed during the preceding 12 months. The annual subscription price for *American Entomologist* is $200 for institutions and $107 for individuals. During the preceding 12 months, the average number of copies printed for each issue was 5454; the average paid print and paid electronic copies circulation 5174; the average free distribution 13; the average number of print and paid electronic copies distributed 5187. Corresponding figures for the last issue before filing: total number of copies printed 5470; total paid circulation 5124; free copies distributed 12; total distribution 5136. Average percent paid 99.75; actual percent paid 99.77.

I certify that the statements made by me above are correct and complete.

Jennifer Melvin
U.S. Journals Subscriptions Supervisor
Oxford University Press