

Defining the experimental unit in grazing trials

Dwight S. Fisher¹

USDA-ARS, JPCS Natural Resource Conservation Center, Watkinsville, GA 30677-2373

Abstract

An experimental unit is the portion of experimental material to which a treatment is applied. Randomization of treatments is implied among experimental units. This simple definition is often overlooked, but statistical analyses are predicated on identification of the experimental unit. Experimental error is the variation among observations of experimental units treated alike. The group of animals on a pasture treatment is a single experimental unit. Redefining the experimental unit as the animal on the pasture rather than the pasture will erroneously increase the degrees of freedom and increase the likelihood of declaring a significant difference, but the test will be invalid. Regression analysis has been used as an alternative to traditional replication, but it results in risks that some scientific communities find unacceptable. Using a continuous variable, each treatment is applied to a single experimental unit (no true replication) and then deviation from a regression model is used to estimate experimental error. This approach has resulted in strong differences of opinion in the analysis and publication of stocking rate studies without replication of treatments. Objective standards are the core of experimental design, but there are still subjective decisions that must be made on the utility of data. The publication of a single replication for later metanalysis or the use of an alternative approach such as deviations from a regression model could be justified in some cases of unique or extensive experimental units.

Key Words: Statistics, Experimental Design

Introduction

An experimental unit is the portion of experimental material to which a treatment is applied (Steel and Torrie, 1980; Gomez and Gomez, 1984). This simple definition is often overlooked or misunderstood. The experimental units used for an experiment should also be representative and random samples of the population to which inferences will be made. However, logistical constraints generally prevent a rigorous application of that rule. Researchers typically use statistical analyses that are predicated upon an accurate identification of the experimental unit. The experimental unit is the foundation of the definition of experimental error, which is the variation among observations of experimental units treated alike. Randomization is an implied characteristic of distributing treatments among experimental units.

Discussion

An Example from Plant Research

If an experiment is conducted without multiple experimental units the observations made during the experiment do not contain the information needed for an estimate of the experimental error. On the other hand, absence of randomization means that the estimate of experimental error, even if there is replication, may be seriously biased (Lee and Rawlings, 1982). Even when using controlled environmental chambers in plant science research, the plants within chambers are subsamples and are not replicates. Lee and Rawlings (1982) examined this issue in depth and found that, even with the most uniform conditions that could be

created using state-of-the-art environmental chambers, the basic concepts of experimental design could not be violated without losing the ability to accurately estimate experimental error. Poor design resulted in poor estimates of error. This can result in the declaration of a significant difference when in reality there is none, and of course the estimates of probability associated with the estimated differences are erroneous.

When subsamples are incorrectly represented as replications an unknown level of bias is introduced in the estimation of treatment effects and the experimental error is underestimated. Even in a uniform environment, grouping of the experimental material for a given treatment and handling the group as a unit introduces positive correlations within the group, and the net result is that the experimental error is underestimated if subsamples are mistakenly identified as replications (Lee and Rawlings, 1982).

Rangeland, Pasture, and Forage Research

When studying animal performance in grazing systems, there are two primary sources of variation within each year (Mott and Lucas, 1953): the variation as a result of pasture or rangeland treatment and the variation due to animals. Generally, animal-to-animal variation is thought to be two to six times as large as paddock-to-paddock variation (Mott and Lucas, 1953). In grazing research, the group of animals on a pasture generally makes up a single experimental unit (Stuedemann and Matches, 1989). The exception to this rule is when a treatment such as a drug treatment is applied as a split plot to one or more animals in an individual paddock and one or more animals in the same paddock receive a different treatment. Split-plot designs actually have more

than one estimate of experimental error, and it varies because there are, by definition, multiple sizes or types of experimental units within the design.

When the experimental material is made up of discrete units such as cattle, there can be a tendency to assume that the unit of experimental material is the same as the experimental unit. When treatments are applied to an entire pasture (i.e., a forage species or herbage mass treatment), each animal on a pasture is a subsample and the group of animals is the experimental unit. The complex logistics and expense of conducting a grazing trial can produce a strong desire to redefine the experimental unit when more than one animal grazes each pasture. This is because of the degrees of freedom in the analysis of variance in most grazing experiments. However, few scientists would argue that the animals on a pasture have no influence on each other, are not affected similarly by a common environment, and as a result do not exhibit responses that are correlated.

When the experimental unit is misidentified as a subsample, the correlation of subsamples within an experimental unit results in an erroneous estimate of experimental error that is too small. As a result, it can seem advantageous to redefine animals as replications in the analysis of variance of grazing trials simply because it makes it easier to find a “significant” difference. This is at the root of the problem identified by Lee and Rawlings (1982) and it is much more serious in the complex experiments that make up grazing trials in both pasture and rangeland. Any flaw in experimental design that allows for correlation of animal responses within a treatment across replications results in reduced estimates of experimental error and increases the possibility of declaring a significant difference when there is none. The experimenter will minimize animal-to-animal variation by placing animals together in a paddock to be treated similarly (Mott and Lucas, 1953). Even if no treatments were falsely declared to be different, the probabilities associated with the differences would be in error. This is a product of misidentification of the experimental unit and consequently using an inappropriate statistical model.

A Simple Example with Animals Fed in Confinement

In the case of a simple trial to test three hays (Figure 1), a researcher may acquire the hays and then randomize the feeding to each animal. For example, consider the use of six animals as shown in Figure 1. As indicated, the hays are randomized to each of six animals and, in the example, three time periods are used to feed each animal each hay and form two Latin squares. This design would generally be publishable in the *Journal of Animal Science* if we assume that the treatments selected make a contribution to the scientific literature. Multiple environments and a larger number of animals would be desirable but not necessary. In contrast, if the same hays were fed for the same time period with animals in separate pens but grouped by treatment to simplify feeding (Figure 2), most animal scientists would protest that this design was inappropriate, despite the fact

that the same numbers of animals were fed for the same duration.

Although the treatments may have been assigned at random to the pairs of animals, each pair was fed the same hay side by side. Even if the animals are in separate pens, grouping the feeding treatments tends to correlate the response variables of the animals receiving the same treatment side by side. It should be noted that the trial represented in Figure 2 could be turned into a Latin square with each pair of animals making a cell if two additional periods were added. However, that is not proposed here, and it would triple the duration of the trial. Assume simply that the animals are in separate pens but that the feed treatments are grouped for logistical reasons and no other animals are used to achieve replication.

In the proposed example (Figure 2) the pair is the experimental unit and there is no replication. Most animal scientists would object to the flawed design, point out the lack of replication, and recommend that the manuscript not be published. This design is analogous to a single replication of a typical grazing trial. Animals are grouped and actually penned in the same paddock for a pasture treatment. The group is the unit because they are all together feeding on the same treatment. Conceptually, that portion of the experimental material is not divided and forms a unit. They are expected to display correlated responses to any treatment variables. Most animal scientists would have natural and appropriate objections to grouping animals by treatment in any trial unless the trial was designed to use groups as experimental units. Most grazing trials are designed to use the group as the experimental unit. They use the group as an experimental unit and replicate groups with multiple pastures.

Exceptions to the Rule and Lack of Replication

The first exception to optimal experimental design I would like to point out is illustrated in Figure 1. Notice that only three hay fields are used, each producing a single experimental hay. This is generally the way experimental hay treatments are generated. However, this means that any peculiarity of the field used to produce the hay is confounded with the experimental treatment. The desired inference is generally to all hays of that type at a similar stage of growth, and this could result in serious errors. This is generally acceptable because of the logistics of replicating hay sources and incorporating them into the experimental design. The *Journal of Animal Science* has established a de facto standard that permits this exception to what would otherwise be a more robust experimental design.

The primary topic I need to address in this section is the use of regression analysis in grazing research as an alternative to replicating treatments with experimental units. This approach uses deviation from regression to estimate experimental error. Without multiple experimental units for each treatment some scientific communities find these designs unacceptable. For example, Hart and Ashby (1998) published a grazing experiment covering 55 yr of research

in the *Journal of Range Management*. Although few would question the utility of these data, it is unlikely that this experiment would have been published in the *Journal of Animal Science*, *Crop Science*, or *Agronomy Journal* because it lacked replication of each treatment with multiple experimental units and error was estimated solely by deviation from the regression model.

Some scientists review manuscripts as though the use of regression analysis, rather than replication of the experimental unit, is not statistically valid. The use of regression with multiple levels of a treatment effect to predict a response variable is statistically valid but in some scientific communities it may not be sufficient for publication. In this approach, multiple experimental units are used but generally each treatment is applied to only one experimental unit. When a regression model is used to estimate error the estimate should actually be a conservative estimate. For example, although we may use a model that provides an adequate fit to observations the model will be a gross simplification of the "true" model. For example, a linear or quadratic model is sometimes applied to explain a relationship because the models are simple but the relationship under study is not thought to be truly linear or quadratic. As a result, deviations from the models should be larger and the error term inflated because we are not using the "true" model. In grazing research, this issue has been most evident in the analysis and publication of stocking rate studies without replicated experimental units.

Differences in the definition of acceptable designs may be partly responsible for the diverse theories on the relationship of stocking rate to ADG in the literature and for the association of these theories with particular regions and environments (Hart, 1993). Using varied stocking rate and deviations from regression analysis to estimate experimental error is sometimes associated with risks (real and perceived) of inadequate estimation of experimental error and model parameters. The relative value of eliminating replication in favor of multiple stocking rates in some grazing experiments has been considered (Riewe, 1961; Bransby et al., 1988; Bransby, 1989).

I have presented a hypothetical result from a grazing trial without replication in a diagram (Figure 3). The graph represents observations of ADG at three stocking rates. One serious problem with the approach is evident with this minimum implementation of only three levels of treatment (stocking rate). In effect, the slope is determined by the observations of ADG at the maximum and minimum rates of treatment, whereas the error is determined by the deviation of the observed ADG from the regression line for the intermediate treatment. The situation is not as serious when more than three levels of treatment are applied. However, influential points representing single pastures are always possible issues with this approach. It should also be noted that even though the error estimated for the slope is limited it could be used to apply a level of confidence to both the slope and the estimated optimum gain per hectare.

When replication of the experimental unit would prevent the investigation of a scientific principle, the use of an

alternative design clearly should be considered. For example, scale can be of importance when evaluating management systems and in interpretation of grazing data (Hart et al., 1993). Particularly as we move up to the analysis of systems at full scale we do not have the option of simply saying we cannot do the work. In some cases it is appropriate to publish a single replication in anticipation that other researchers will publish another. This is handled well in the *Journal of Range Management* and the *Journal of Environmental Quality*. In both of these journals, variation in the size or unique characteristics of the experimental unit is taken into account, and great diversity is seen within an individual volume of these journals.

Each scientific community interacts to establish de facto standards for experimental design and analysis. Studies conducted on experimental units that are very large or unique may have no replication, but it may be appropriate to publish a single replication as a case study for comparison with other similar case studies. In part, standards are based on the logistics of conducting research in a particular field of science and an acceptable level of risk in the results. Less rigorous designs may be appropriate for conducting experiments that are very difficult, expensive, or extensive.

Implications

The logistics of grazing research typically result in experimental designs with limited degrees of freedom for estimating treatment differences. This makes it difficult to detect significant effects, but it is inappropriate to use observations of each grazing animal as replications when the treatment is applied to the pasture. This practice can erroneously increase the degrees of freedom, but it is inappropriate to assume that the experimental unit can be redefined. On the other hand, scientific societies may establish rigid standards that exclude data that could be of value. Objective standards are the core of experimental design, but there are still subjective decisions that must be made on the utility of data. The publication of a single replication for later metanalysis or the use of an alternative approach such as deviations from a regression model could be justified in some cases of unique or extensive experimental units.

Literature Cited

- Bransby, D. I. 1989. Compromises in the design and conduct of grazing experiments. p. 53-67. In: G.C. Marten (ed.), *Grazing Research: Design, Methodology, and Analysis*. Crop Science Society of America. Special Publication no. 16. Madison, WI.
- Bransby, D. I., B. E. Conrad, H. M. Dicks, and J. W. Drane. 1988. Justification for grazing intensity experiments: Analysing and interpreting grazing data. *J. Range Manage.* 41:274-279.
- Gomez, K. A., and A. A. Gomez. 1984. *Statistical Procedures for Agricultural Research*. 2nd ed. John Wiley, New York.
- Hart, R. H. 1992. Viewpoint: "Invisible colleges" and citation clusters in stocking rate research. *J. Range Manage.* 46:378-382.

- Hart, R. H., and M. H. Ashby. 1998. Grazing intensities, vegetation, and heifer gains: 55 years on shortgrass. *J. Range Manage.* 51:392-398.
- Hart, R. H., J. Bissio, M. J. Samuel, and J. W. Waggoner, Jr. 1993. Grazing systems, pasture size, and cattle grazing behavior, distribution and gains. *J. Range Manage.* 46:81-87.
- Lee, C., and J. O. Rawlings. 1982. Design of experiments in growth chambers – Uniformity trials in the North Carolina State University Phytotron. *Crop Sci.* 22:551-558.
- Mott, G. O., and H. L. Lucas. 1953. The design, conduct, and interpretation of grazing trials on cultivated and improved pastures. In: *Proc. 6th Int. Grassl. Congr., Aug. 1952, Pennsylvania State Univ., State College, PA.* pp 1380-1385.
- Riewe, M. E. 1961. Use of the relationship of stocking rate to gain of cattle in an experimental design for grazing trials. *Agron. J.* 53:309-313.
- Steel, R. G. D., and J. H. Torrie. 1980. *Principals and Procedures of Statistics (2nd Ed.)*. McGraw-Hill, New York.
- Stuedemann, J. S., and A. G. Matches. 1989. Measurement of animal response in grazing research. In: G. C. Marten (Ed.) *Grazing Research: Design, Methodology, and Analysis*. Special Publication no. 16. pp 21-35. Crop Science Society of America, Madison, WI.

Notes

1. Correspondence: 1420 Experiment Station Road, (phone: (706) 769-5631, ext. 268; fax: (706) 769-8962; E-mail: Dwight.Fisher@Scientist.com).

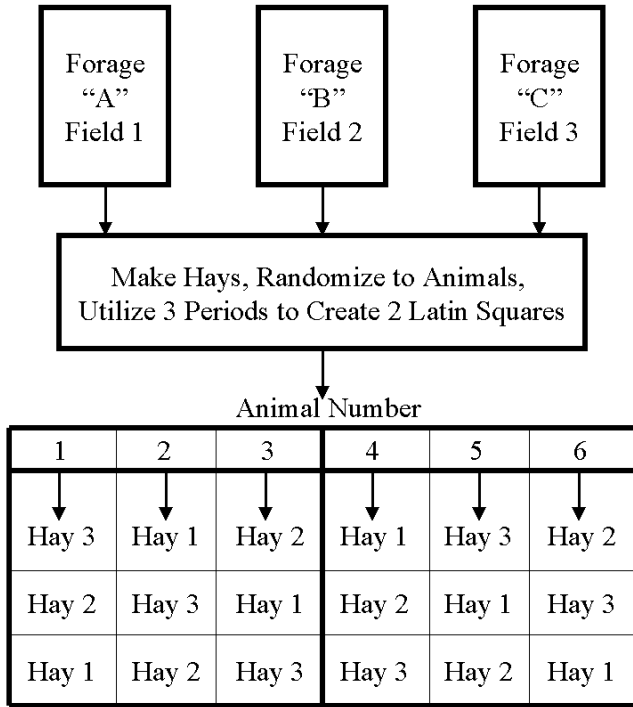


Figure 1. Diagrammatic representation of a double Latin square design in a feeding trial.

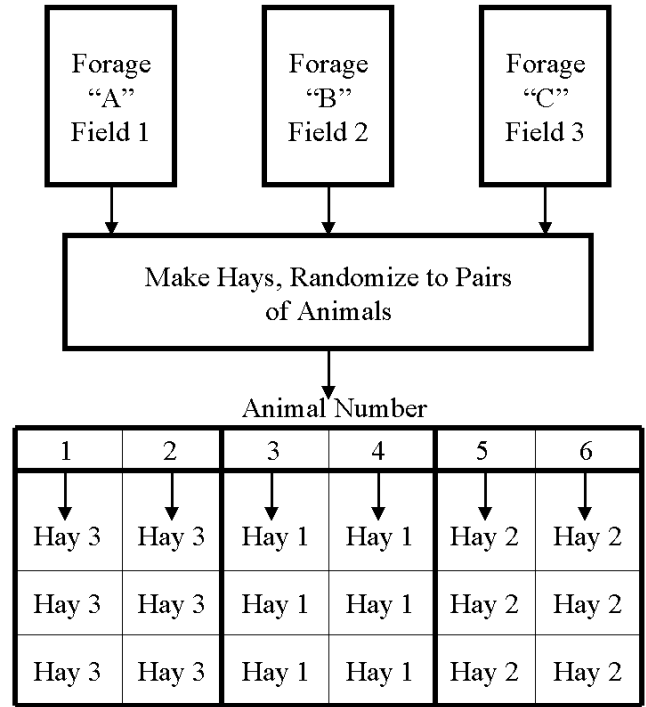


Figure 2. Diagrammatic representation of a feeding trial with feed treatments grouped to make for simpler logistics.

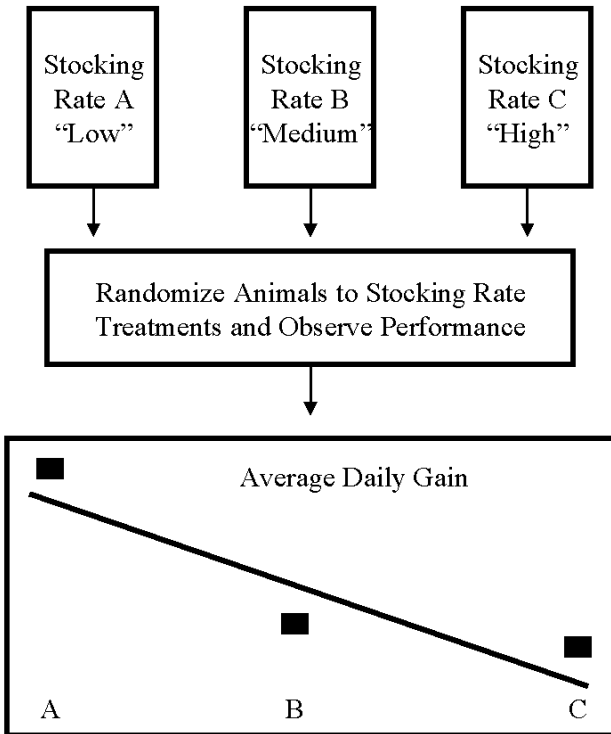


Figure 3. Hypothetical average daily gains in a stocking rate study with three rates.