PSEUDOREPLICATION – PART 1

By David B. South and Greg L. Somers - School of Forestry and Wildlife Sciences,

Auburn University

This essay is for those who apply treatments (e.g. herbicides, fire, cultivation, genotypes, fertilizer, cover-crops, organic matter, etc.) in order to test hypotheses and for those who rely on this research when making operational choices. It is also directed at those who reject the null hypothesis after taking samples taken from different natural populations. It is not directed at those who sample from natural populations and then formulate hypotheses (ie. no conclusions are presented).

Professional foresters need sound scientific information to determine what stock-type to plant, when to thin and how much fertilizer to apply. Unfortunately, some researchers establish non-replicated trials and report findings that are based on pseudoreplication. Those who use pseudoreplication in order to reject a null hypothesis run a risk of making a Type 1 statistical error (ie. claiming a treatment effect is real when it is not).

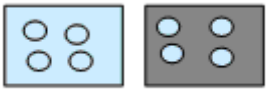
In some fields, more than 1 out of 4 papers involve pseudoreplication (Hurlbert 1984; Heffner et al. 1996). Although some graduate students are exposed to the concept of pseudoreplication (Kolb et al. 2001), others are not so fortunate. The objective of this essay (the first part of a two-part essay) is to shine some light on this topic.

Pseudoreplication occurs when treatments are not replicated but a t-test or an ANOVA (analysis of variance) is carried out anyway by assuming sub-samples (or in some cases individual trees) are the same as replication. This results in use of an incorrect number of error degrees of freedom. A major factor that leads to pseudoreplication is a failure to define the correct experimental unit. In addition, some journal editors do not require ANOVA tables be included in manuscripts. As a result, a reviewer might not know if the error term involved 13 or 2841 degrees of freedom. Reviewers who would normally reject non-replicated trials can sometimes be fooled into thinking that a statistical analysis (that involves pseudoreplication) is valid.

When probability values are published using pseudoreplication, the values are suspect since they might claim a significant difference where none really exists. Let us examine a “real world” case where five treatments were replicated 5 times (producing 25 completely random experimental units with 24 measured trees per experimental unit). In this case, the correct error term has 20 degrees of freedom and there is no significant treatment effect (P= 0.253). However, if pseudoreplication occurs (producing 50 pseudo-experimental units, each with 12 trees per mean) the error degrees of freedom increases to 45 and the treatment then becomes “magically” significant (P= 0.0426). An even lower value (P=0.0023) is obtained when 100 pseudo-experimental units are created (each with 6 trees per mean). Therefore, the experimenter can reduce the P-value simply by changing the number of trees in each pseudoreplication.

According to Hurlbert (1984), there are several types of pseudoreplication: simple, sacrificial, temporal and implicit. We describe five types of pseudoreplication.

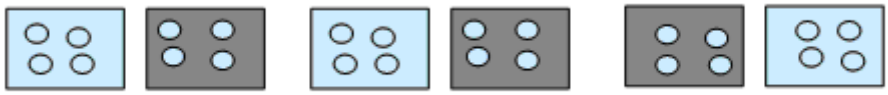
(1) Statistical tests involving subsampling and no replication (a.k.a. simple pseudoreplication)



This example has 4 pseudoreplications (circles) per experimental unit (squares) with no treatment replication. The light box is the "control" treatment. A proper ANOVA is not possible since there are no true replications.

Of course statistics can be run on just about any set of data. A t-test can be run on subsamples from the control (21,20,17,22) and treated (23, 22, 23, 24) units above. Such a test would show a significant difference in means (20 vs 23) with an LSD (Least Significant Difference) value of 2.8. The probability of a greater t-value would be P=0.0408. It would be correct to say there is a significant difference in the two populations, but incorrect to claim the treatment caused the 15% increase. The 15% increase might be due to shade covering the control plot during part of the day, or from soil in the treated plot having a higher level of fertility. Researchers should not simply assume there are no confounding factors associated with plot location. For example, Easley (1963) assumed that plot location on a sandy site (or on a clay site) would have no effect on growth of pines after 5 years in the field. In addition, he assumed that nursery management practice would have no effect on early growth so he compared "nursery run" seedlings grown at a state nursery with "seed-production-area" seedlings grown at an industrial nursery. Although he had no replications of seed source in a nursery (or on either field site), he concluded that the collection of seed from a local seed production area is worth the investment of time and money. Forty years later, some journal authors still report results from no replications. Some use pseudoreplication to test for significant differences among treatments (e.g. Younger et al. 2008).

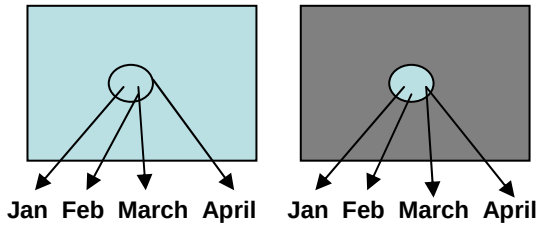
(2) Statistical tests assuming subsamples are experimental units (a.k.a. sacrificial pseudoreplication)



This example of a completely randomized design has 4 pseudoreplications per experimental unit with 3 replications. The correct ANOVA to test for a treatment effect would have an error term with only 4 degrees of freedom (using one mean per experimental unit). An incorrect ANOVA would have an error term with 22 degrees of freedom. An example of a paper that had more error degrees of freedom than experimental units was published in 2009 (Verble and Stephen 2009).

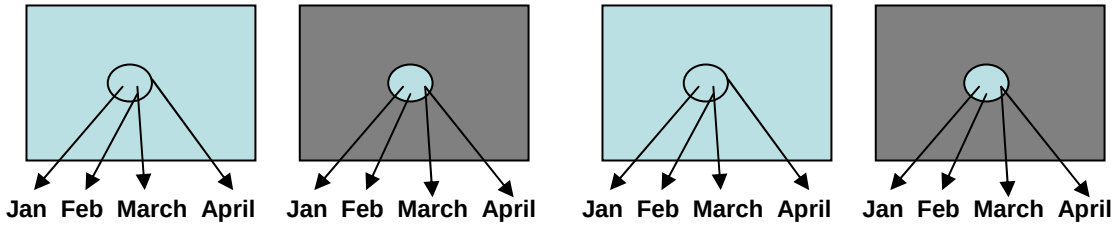
When an incorrect ANOVA is applied to the example above, and the null hypothesis is not rejected, a Type 1 error does not occur. However, if the null hypothesis is rejected with the incorrect ANOVA, then readers may question the conclusion since the authors might have increased their chance of making a Type 1 error.

(3) Statistical tests with no replication but assuming sampling over time (a.k.a. temporal pseudoreplication)



This example has 4 pseudoreplications (repeated measures) per experimental unit with no treatment replication. A proper ANOVA is not possible since there are no replications. Some recent papers include sites with pseudoreplication (e.g. Blankenship and Arthur 1999).

(4) Statistical tests involving replications and sampling over time (a.k.a. temporal pseudoreplication)

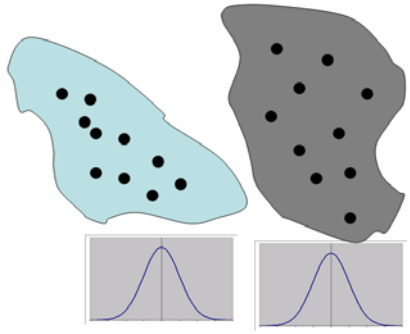


This example of a completely randomized design has 4 pesudoreplications (repeated measures) per experimental unit with 2 treatments and 2 replications. The tree in each experimental unit is measured in January and remeasured in February, March and April. A correct ANOVA might test treatment (1 d.f.) with an error term with only 2 degrees of freedom. It does not matter how many repeated measures are made, the number of experimental units remains the same. The example above only has 4 experimental units and, therefore, an incorrect ANOVA would have an error term with 14 degrees of freedom.

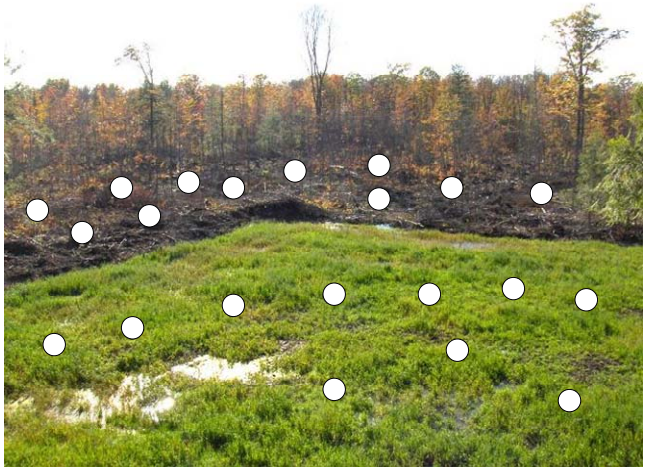
There are many correct statistical procedures to properly analyze replicated experiments with multiple measurements of each experimental unit collected over time. The proper procedure for a given experiment depends upon the nature of the measurements over time and the hypotheses that the experimenter wants to investigate. If there is no reason to suspect any time differences then these measurement are analogous to subsamples of the experimental unit and simple averaging over time would be the best approach. For experiments where there are differences over time that affect all experimental units, then multivariate procedures can be used to account for these differences without having to specify how the values change over time. These procedures provide statistical tests of not only the treatments, but also the treatment by time interaction which tells the researcher whether the treatment effect is the same at all time periods.

A lot of experiments in forestry deal with trees and the growth of trees over time. In these cases, the final measurement is the accumulation of the treatment effects over the course of the experiment, which means a standard ANOVA can be used. The time trends can also be shown in graphs. If these time trends are also of interest then a growth function can be fit to each experimental unit to describe the change over time. The parameters themselves can then be tested for equality using multivariate procedures or a single function of the parameters, such as the maximum or asymptotic value, could be tested in a standard ANOVA. This approach can often lead to a more in depth understanding of how the treatments affect the experimental units over time than a simple test of the interaction between treatments and time. Additional information on details of appropriate repeated measures analyses include: Moser et al. 1990; Meredith and Stehman 1991; Gumpertz and Brownie 1993; Zhao et al. 2005.

(5) Subsampling with no statistical tests (a.k.a. implicit pseudoreplication)



This example is similar to case #1 except here no statistical test is made. This example has 10 samples per site and these samples are used to estimate a normal distribution. However, Hurlbert (1984) said it is "useless" to discuss "differences" between the two treatments by comparing the overlapping distributions. It does not matter if a statistical test is carried out, it would be incorrect to reject a null hypothesis based only on samples taken from just two areas. Unfortunately, several examples of "implicit pseudoreplication" can be found in the forestry literature (e.g. Dong and Burdett 1986; Kamaluddin et al. 2005). Due to potential confounding of other factors, one should avoid drawing conclusions about "treatments" (e.g. burn; agriculture; fertilization; etc.) when there is no true replication of experimental units.



Dots represent soil samples taken from a burned forested area and an adjacent unburned pasture. Conclusions about burning or agriculture should not be made by comparing the means or distributions from these samples since these two areas might have inherently different soils. Unfortunately, some authors of papers incorrectly make conclusions about factors that are based just on a comparison of means of subsamples taken from two separate locations.

Foresters should realize that researchers differ in their opinions regarding what constitutes "scientific proof." Some require replication of treatments while others make inferences about treatments that are based solely on pseudoreplication.

REFERNCES

Blankenship, B.A. and M.A. Arthur. 1999. Prescribed fire affects eastern white pine recruitment and survival on eastern Kentucky ridgetops. Southern Journal of Applied Forestry 23(3): 144-150.

Dong, H. and A.N. Burdett. 1986. Chemical root pruning of Chinese pine seedlings raised in cupric sulfide impregnated paper containers. New Forests 1:67-73

Easley, L.T. 1963. Growth of loblolly pine from seed produced in a seed production area vs. nursery run stock. Journal of Forestry 61(5):388-389.

Gumpertz, M.L. and C. Brownie. 1993. Repeated measures in randomized block and split-plot experiments. Canadian Journal of Forest Resources 23:625-639.

Hurlbert, S.H. 1984. Pseudoreplication and the design of ecological field experiments. Ecological Monographs 54:187-211.

Heffner, R.A., M.J. Butler, IV. and C.K. Reilly. 1996. Pseudoreplication revisited. Ecology 77(8): 2558-2562.

Kamaluddin, M.S., X.Chang, M.P. Curran and J.J. Zwiazek. 2005. Soil compaction and forest floor removal affect early growth and physiology of lodgepole pine and Douglas-fir in British Columbia. Forest Science 51(6):513-521.

Kolb, T.E., P.J. Daugherty and L.E. DeWald. 2001. Paving the way from school to work. Journal of Forestry 99(9):10-15.

Meredith, M.P. and S.V. Stehman. 1991. Repeated measures experiments in forestry: focus on analysis of response curves. Canadian Journal of Forest Research 21(7):957-965.

Moser, E.B., A.M. Saxton and S.R. Pezehski. 1990. Repeated measures analysis of variance: application to tree research. Canadian Journal of Forest Research 20:524-535.

Verble, R.M. and F.M. Stephen. 2009. Occurrence of carpenter ants in Ozark forests in relation to prescribed fire and stand variables. Southern Journal of Applied Forestry 33(1)42-45.

Younger, N.L., L.H. Temesgen and S.M. Garber. 2008. Taper and volume responses of Douglas-fir to sulfur treatments for control of Swiss needle cast in the Coast Range of Oregon. Western Journal of Applied Forestry 23(3):142-148.

Zhao, D., M. Wilson and B.E. Borders. 2005. Modeling response curves and testing treatment effects in repeated measures experiments: a multilevel nonlinear mixed effects model approach. Canadian Journal of Forest Research 35:122-132.