Design and Analysis of Commonly Used Comparative Horticultural Experiments

George C.J. Fernandez[1]

College of Agriculture, Biotechnology and Natural Resources, University of Nevada–Reno, Reno, NV 89557

A comparative horticultural experiment is a procedure for collecting scientific data in a systematic way to maximize the chance of testing a research hypothesis correctly. It is very important that comparative horticultural experiments are well designed, correctly analyzed, and reported accurately to achieve the full potential of the research. Horticulturists are strongly encouraged to review standard experimental designs statistics textbooks (Kuehl, 1996; Littell et al., 2006), online statistical resources (Government of Alberta 2006; Schabenberger 2006; Sit,1995), and several published articles on advanced statistical methods related to designed experiments (Chew, 1976; Jones, 1984; Petersen, 1977; Piepho et al., 2003; Saville, 1990; Waldo, 1976) before designing their comparative studies and analyzing their data. Also, it is wise always to consult with experimental station statisticians 1) before designing comparative studies, 2) during the analysis, and 3) when interpreting the results. It is not my goal to discuss in this paper an extensive description of theory (this is beyond the scope of this paper), flawed and proper examples, or a description of statistical methods applicable to all comparative horticultural experiments. The aim of this paper is to document the checklists and statistical guidelines applicable to commonly used horticultural experiments, which I presented during a colloquium at the 2004 American Society for Horticultural Science conference in Austin, TX. Download instructions are given in this paper to obtain a hardcopy of this PowerPoint presentation, which contains additional information on sample SAS codes for analyzing data from comparative horticultural experiments discussed in this paper. By reviewing these recommended statistical texts, online resources, and journal papers, and by following the guidelines presented herein, horticulturists may ensure that their research publications conform to the very highest scientific standards with respect to experimental design and statistical methods.

STATISTICAL ISSUES RELATED TO DESIGNING HORTICULTURAL EXPERIMENTS

Designed experiments are considered randomized, controlled comparative studies containing adequate numbers of true replications. A well-designed experiment will avoid bias and will be sufficiently powerful to detect treatment effects likely to be of biological significance. Adequate number of replications is necessary to estimate the experimental error and to have enough power to detect true treatment effects. Proper randomization ensures unbiased estimation of treatment effects and experimental errors. Four types of designed experiments are common in horticultural research: 1) field experiments, 2) greenhouse experiments, 3) growth chambers studies, and 4) laboratory experiments. The characteristics of these different experiments and the important statistical issues related to statistical designs are described next.

Field experiments. Field studies are suitable for conducting yield trials, genotype screening, and comparing horticultural management practices, such as irrigation, nutrition, pest, disease, weed control, and so on. The statistical issues related to designing field experiments are noted in the following paragraphs.

Multiyear studies are important to investigate the interaction between the treatment and the year to remove the confounding effects of climatic variation. The year effects are usually treated as "random effect" in the subsequent analysis.

Multisite studies in a given year are important to evaluate the treatment effects across different sites and to investigate the interaction between the treatment and the site, and to remove the confounding effects of soil variation. If the field investigation is aimed at testing specific site effects and making inferences only about these specific sites, then the site effect should be treated as "fixed effect" in the subsequent analysis. The site effects should be treated as "random effect" when the selected sites represent a random selection of potential sites and the treatment comparison is aimed at making a nonsite-specific overall treatment comparison.

Individual homogeneous field plots with more than single plants are commonly used as experimental units in annual and perennial crop studies. Blocking across the field variability gradient or measuring additional continuous covariates from each experimental unit and using analysis of covariance (Fernandez, 1990) are highly recommended in reducing experimental errors in designed comparative experiments. In tree crop studies, individual trees are usually treated as the experimental units. In both cases, the treatments should be randomly assigned to the experimental units and each experimental unit should have an equal chance of receiving one of the treatments. All response measurements, such as yield, height, biomass, and other horticultural attributes should be recorded for each experimental unit and used in the analysis. Recording response measurements from subsamples within each experimental unit adds additional source of subsample variation and provides a means to estimate both experimental and subsample error that complicate the data analysis of horticultural experiments with subsamples. Therefore, when there is no missing value in the data, multiple subsample measurements taken within an experimental unit (individual plots or trees) could be averaged to simplify the analysis. However, to estimate both the experimental error variation and the subsample error variation, and to use the proper error term to test the treatment effects, a mixed model analysis should be used.

Repeated response measurements are usually taken from the same experimental unit in tree and perennial crops in multiple years or from multiple harvests within the same year and when both are in consideration (e.g., with multiple year trial with many harvests each year). Because these repeated measurements are recorded within the same experimental units, these measurements are considered nonindependent and underestimate the experiment errors. Therefore, repeated-measures analysis of variance (ANOVA) (Fernandez 1991; Littell et al., 2006) should be used when analyzing the horticultural experiments involving repeated measurements to test the significance of the repeated-measures factor and any interaction effects associated with the repeated-measures factor.

Greenhouse or glass house experiments. Greenhouse experiments for treatment comparisons are usually conducted under managed environments using artificial growth media and controlled irrigation and atmospheric conditions to minimize soil and climatic variation. The commonly used experimental units are single or multiple plants grown in individual pots or trays. Although greenhouse experiments are conducted under managed or controlled environments, air temperature, humidity, solar radiation, and moisture variation may occur across the greenhouse as well as within the greenhouse benches. To control the experimental error resulting from the environmental variations, suitable blocking and the use of additional covariates must be used. Furthermore, if the experimental units are pots or planting trays, these could be randomly rotated within a bench during the experimental period to minimize any systematic variation and to reduce edge effect in the greenhouse. The statistical issues related to analyzing greenhouse experimental data are noted in the following paragraphs.

[1]To whom reprint requests should be addressed; e-mail gcjf@unr.edu

Multiyear or multisite studies are uncommon because both climatic and soil variations are controlled here. Main emphasis is given here to compare the main effects and interaction among treatment combinations.

Individual homogeneous pots or trays with more than single plants are commonly used as experimental units in greenhouse studies. The treatments are randomly assigned to the experimental units (individual pots or trays with more than single plants), and each experimental unit should have an equal chance of receiving one of the treatment combinations. Recording response measurements from subsamples (different plant parts within a single plant or pot) within each experimental unit is also very common and adds additional source of subsample variation. Therefore, multiple subsample measurements taken within an experimental unit (individual pots) should be averaged when there are no missing values to simplify the analysis or a mixed model analysis should be used to separate the experimental error variation and the subsample error variation.

Repeated response measurements are also very common in greenhouse studies, and repeated-measures ANOVA should be used when analyzing the horticultural experiments involving repeated measurements.

Growth chamber experiments. Growth chamber studies are special controlled environment studies in which the researcher can study the temperature or light effects on plant growth by controlling specific environmental factors at a given level throughout the experimental study. The lack of or inadequate number of replications is the main problem in growth chamber studies. Because the treatment condition (fixed temperature or light intensity/duration) are randomly applied to the growth chamber and not to the pots or individual plant, the growth chamber unit becomes the experimental unit. The pots, Petri dishes, trays, and plants included within a growth chamber are considered subsamples and should not be treated as true replicates. To provide valid replications when the number of growth chambers are limited, the experiments should be repeated by using the same growth chamber with different randomly assigned treatment levels multiple times to collect statistically valid experimental data.

Laboratory experiments. Horticultural laboratory experiments can be classified into two major categories. The first is studies involving chemical or nutritional analysis of plant samples obtained from other field or greenhouse studies. These are extensions of the field or greenhouse studies, and therefore the experimental design and same sampling plans should be used in the statistical analysis of the laboratory investigations. The second category includes stand-alone experiments using horticultural materials (e.g., seeds, tissue culture propagules, cellular components such as proteins, fats, carbohydrates, hormones, postharvest plants, and plant products) handled under laboratory conditions.

Thus, the recommended statistical principles in conducting valid experiments such as randomization, adequate replication, and error control measures should be used in designing these laboratory experiments involving plant and their parts.

STATISTICAL ISSUES RELATED TO ANALYZING DATA FROM HORTICULTURAL EXPERIMENTS

The data collected from comparative experiments should be evaluated using the appropriate statistical analysis. The aim of the proper statistical analysis is to extract all valid information present in the data in such a way that it can be interpreted, taking into account the inherent variability and measurement errors associated with the experiment. The choice of statistical analysis will depend on the purpose of the study, the design of the experiment, and the nature of the response variable.

Designed experiments conducted by horticultural scientists often give rise to several random sources of variation. Relevant examples are split plot designs, multiyear and multisite yield trials, and repeated measurements taken on the same field plot. Unbalanced data are a common problem in plant research based on designed experiments. The analysis of repeated measurements needs to account for the correlation among observations on the same experimental unit. There are many different types of models that can be imposed on the correlation structure, including autoregressive and spatial models. These correlation structures can be embedded in a mixed model framework. Data arising from such experiments may be conveniently analyzed by mixed models.

A fixed effect is considered repeatable (i.e., the same levels of the factor could be repeated by another scientist), and scientists are only interested in making treatment comparisons about the specific levels used in that experiment. A random factor is an effect for which the levels of the factor are considered to be a random sample from an infinite set of possible levels. When the levels of a treatment cannot be exactly replicated by another researcher, then that treatment should be thought of as a random effect. However, even treatments that are repeatable may be considered random effects if they can be regarded as a random sample from some larger population of treatment levels. With random effects, scientists are only interested in estimating the population variances related to randomly selected factor levels. Also, inferences about the random effect are not concerns about treatment mean comparisons (as they are with fixed effects), but concerns about the estimation of the variance covariances of the random effects.

A split plot experiment requires two random error terms for main plots and subplots. Repeated measurements taken at different points in time or space are correlated, which may be accounted for by a mixed model with an appropriate variance–covariance structure. Experiments replicated at several sites or in several years require a linear model with random site or random year effects, providing environments may be considered as a random factor and the objective is to compute treatment means across a wide range of environments.

The mixed procedure of SAS software version 9.1 of the SAS System for Windows (Littell et al., 2006), allows for flexible inclusion of a combination of continuous and categorical random effects as well as nested and hierarchical structures. This will provide restricted maximum likelihood (REML) estimates of variance and covariance components among model factors, and permits fitting both fixed and random model effects in mixed models analyses (Littell et al., 2006).

The statistical guidelines and checklists for analyzing data obtained from commonly used comparative designed experiments involving a continuous response variable are presented in Tables 1 through 4. The choice of the correct statistical analysis depends on the scale of the response variable, type of experimental design [complete randomized design (CRD), randomized complete-block design (RCBD), split plot, and so on], treatment design (single factor, factorial design, fixed vs. random factor, qualitative vs. quantitative factors, independent vs. correlated or repeated factor levels, presence of subsamples, and so on). Therefore, horticulturists are encouraged to review the checklists and statistical analysis guidelines presented in Tables 1 through 4 before analyzing their data and to select the correct analysis plans, including the appropriate mean comparison methods. This exercise will also help them to describe the experimental design used and statistical methods used effectively in the materials and methods section. For additional information regarding running these statistical analyses using SAS software and to see the sample SAS code, please visit my home page at www.cabnr.unr.edu/gf and download a copy of the PowerPoint presentation by clicking the running puppy clip art and following the online instructions.

CHECKING FOR VIOLATIONS OF ANOVA ASSUMPTIONS

Checking for any violations of statistical assumptions such as 1) nonindependence of error, 2) unequal error variation within treatment, and 3) nonnormal error distribution is an important component of statistical analysis. Proper randomization methods in experimental designs ensure that the residuals are independent. However, when randomization is not possible with repeated-measures treatment, experimental errors become correlated, resulting in the underestimation of experimental errors. Therefore, repeated-measures ANOVA using the mixed model methodology should be used to correct this problem. Constant or equal variance assumption can be assessed graphically by plotting the residuals