

Beyond Cowboy Science: Simple Methods for Conducting Credible and Valid Research

Diane L Haase

Diane L Haase, Western Nursery Specialist, USDA Forest Service, PO Box 3623, Portland, OR 97208; E-mail: dlhaase@fs.fed.us

Haase DL. 2014. Beyond Cowboy Science: Simple Methods for Conducting Credible and Valid Research. In: Wilkinson KM, Haase DL, Pinto JR, technical coordinators. National Proceedings: Forest and Conservation Nursery Associations—2013. Fort Collins (CO): USDA Forest Service, Rocky Mountain Research Station. Proceedings RMRS-P-72. 66-77. Available at: http://www.fs.fed.us/rm/pubs/rmrs_p072.html

Abstract: Many nursery and field trials are conducted every year to test new products and techniques. Some of these trials, however, can produce data that is too variable or confounded to accurately assess the question(s) of interest. A “cowboy science” approach can yield results that are statistically invalid and/or biologically untrue; using such data can lead to erroneous conclusions. By incorporating a few simple, basic principles of study design and data collection, anyone can yield credible data that can be used to answer questions or make decisions. Despite beliefs to the contrary, using a valid experimental design usually requires little or no additional input of time and resources, nor does it require an in-depth understanding of statistics. Good research design also ensures that the time and resources invested in research yields meaningful results. This paper describes the “Three Rs” of study design – Representation, Replication, and Randomization – along with examples of pitfalls and successes. It also describes how to create a study plan to guide effective research in the nursery or the field.

Key Words: experimental design, confounding, bias, research methods, data collection.

What is Cowboy Science?

The term “Cowboy Science” was coined many years ago by northwest foresters to refer to “quick and dirty” trials or “demo plots” established operationally to evaluate a technique or treatment (Rose 2000). In no way is this meant to be derogatory to cowboys—quite to the contrary! This term is a nod to the stereotypical cowboy’s independence and resourcefulness in solving problems. Many foresters and other field professionals lack the background or confidence to set up a research project based on statistical theory and design, but most have the intelligence, professional curiosity, and creativity to practice Cowboy Science on occasion. Over the decades, an enormous amount of time, land, and resources have been dedicated to investigating seedling growth in the nursery and after outplanting in response to new products or techniques.

Cowboy Science can be helpful for generating some preliminary observational data used for initial exploration of simple research questions. Such data, however, are considered “anecdotal” and insufficient to adequately or accurately assess the question at hand. Drawing conclusions from such data can be risky.

Risks Associated with Cowboy Science

The inherent characteristic of Cowboy Science is its disregard for experimental methods designed to generate valid data for addressing study objectives. This approach can yield results that are statistically invalid or biologically untrue. Using such data can lead to erroneous conclusions. Using flawed results is especially problematic (and costly) when making management decisions.

Example #1

Cowgirl Jane set up a nursery study to test two products that the manufacturer claims will increase root growth. She applied the products to two nursery beds in an out-of-the-way area of the nursery. Each nursery bed had seedlings from a different low-demand seed lot. She chose these seed lots because she did not want to take the chance of having a negative effect on one of the seed lots she regularly grows in the nursery. She applied Product A to one nursery bed and Product B to the adjacent nursery bed. After several months, she measured 50 of the largest seedlings in each bed and found that those treated with Product B grew more than those treated with Product A. Based on this result, she decided to order Product B for her entire crop. So, what is the problem with Cowgirl Jane's study?

The problem with the study design that Cowgirl Jane used is that conditions in the study area were not uniform. There is a different

seed lot in each nursery bed, and the irrigation patterns result in one bed receiving more water than the other (figure 1). The growth differences she observed could have been due to differences in seed lot or water availability, and therefore have nothing to do with the product she was testing. Additionally, because the treatments were applied to seedlots that are infrequently grown and the study was carried out in an infrequently used area of the nursery, it would be unwise to assume that other seed lots in other areas of the nursery will respond similarly to the treatments. Another issue is that she did not include a control treatment so there is no way to determine if using either of the products results in better or worse root growth than what she does already. Furthermore, data was collected only on the largest seedlings so it is difficult to conclude that the treatment difference is likely to occur throughout the group of seedlings.

Example #2

Cowboy Joe set up a study to compare growth of seedlings from five different nurseries.

He established 5 plots (one per nursery), each with 100 seedlings, on his site. He chose a typical reforestation site to ensure that the study simulated his operational practices. From the onset, he was confident that seedlings from Nursery C or Nursery E would outperform the others. After 3 years, he found that seedlings from nursery C grew the most and decided to sign a large contract with that nursery. So, what is the problem with Cowboy Joe's study?

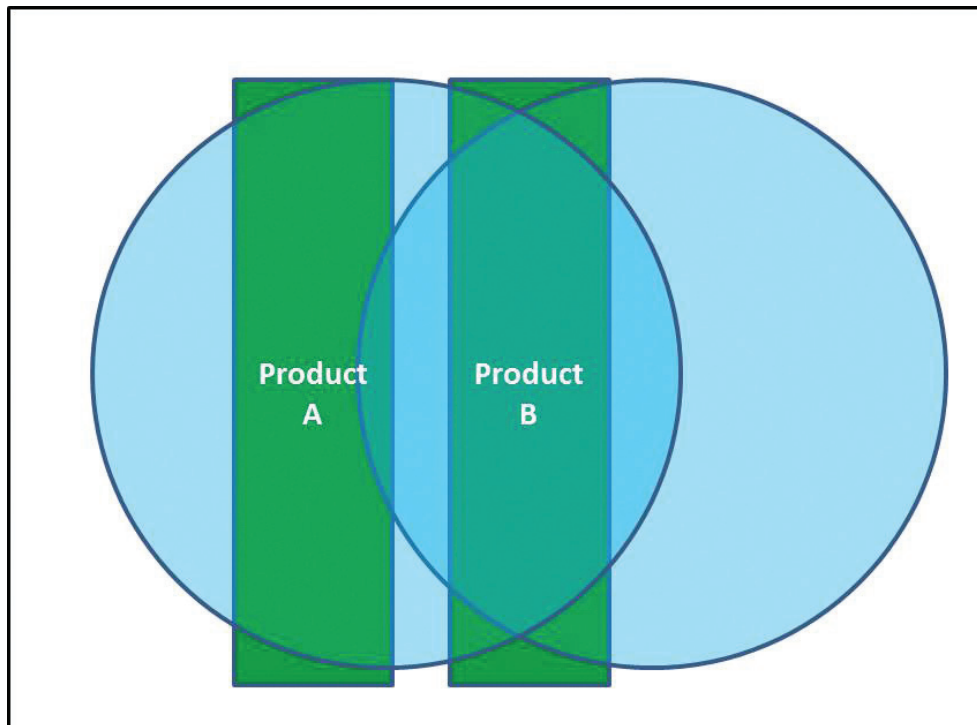


Figure 1. In this Cowboy Science example, a study was installed to compare effects on seedling development of Products A and B applied to two nursery beds. Irrigation patterns, different seed lots, and the lack of a control treatment, however, resulted in confounding and an inability to accurately assess responses to the two products.

The problem with the study design that Cowboy Joe used is similar to the problem with Cowgirl Jane's study design in Example #1—conditions in the study area were not uniform. Because of the variability on the site, conditions in some of Cowboy Joe's plots were more favorable for seedling growth compared to conditions in other plots. Part of the study area was covered with a berry thicket, another part was located where there had been a burn pile, and another part was adjacent to a mature forest resulting in increased browsing and shading (figure 2).

This study design is akin to the adage of having all of one's eggs in one basket—if something goes wrong in one plot, then the study is irreparably compromised. For example, if most of the seedlings in the plot adjacent to the mature forest are severely browsed, then that plot, containing all of the seedlings from one of the nurseries, is effectively eliminated from the study. In addition to the observable variation on the site, there could also be hidden factors such as gradients in soil depth, moisture, fertility, texture, and drainage.

Given the variability on the site, it would be risky for Cowboy Joe to conclude that seedling performance from one nursery is superior to seedlings from other nurseries when, in fact, site conditions may be the primary factor influencing differences in growth and survival among the plots. Furthermore, Cowboy Joe's prejudice in favor of two of the nurseries may have inadvertently swayed the study setup and data collection.

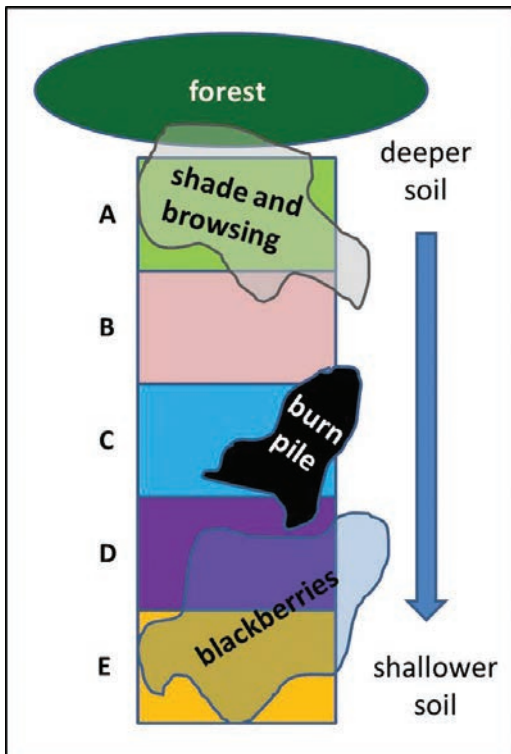


Figure 2. In this Cowboy Science example, five plots were established on a field site to compare seedling growth from five different nurseries (A, B, C, D, and E), but variation in site conditions likely had a greater influence on field performance than the originating nursery. Data from this study design can lead to incorrect conclusions and faulty management decisions.

Confounding and Bias

With regard to study design, confounding and bias can be defined as follows (Dictionary.com 2013a, 2013b):

Confounding—

- to throw into confusion or disorder
- to treat or regard erroneously as identical
- to mix or associate by mistake
- to mingle so that the elements cannot be distinguished or separated

Bias—

- a tendency or inclination, especially one that prevents unprejudiced consideration of a question
- a systematic distortion of a statistic as a result of sampling procedure
- to cause partiality or favoritism in
- to influence, especially unfairly
- selectivity in a sample which influences its distribution and so renders it unable to reflect the desired population parameters

In Cowboy Science, confounding and bias can result in differences among treatments that are not actually due to the treatment. In Example #1, it is impossible to isolate the influences of irrigation pattern, seed lot, and treatment application because those factors are confounded with each other. Furthermore, data in Example #1 that were collected only from the largest seedlings resulted in a biased dataset. In example #2, the effects of nursery source were confounded with the site conditions and the researcher's bias toward the study's outcome may have influenced its design and outcome.

Other Pitfalls of Cowboy Science

In addition to confounding, the Cowboy Science approach often has other aspects that can result in misleading, erroneous, or limiting conclusions. Some of these are:

- **No control treatment** – any study should include a control treatment that allows one to determine how much better (or worse) the new method is compared to the usual way.
- **No study plan** – any study, small or large, needs to have a written plan regarding the objectives, methods, measurements, etc. This plan is important to stay on track and to keep others informed, especially if the person who set up the study is unable to continue it to completion.
- **No labeling or mapping** – it is important to have the study clearly labeled and mapped so that it can be re-visited for future measurements without any questions regarding plot and treatment identification.
- **No follow-through or maintenance** – it is a waste of time and effort to set up a study only to abandon it later due to changes in personnel, poor time management, lack of documentation, or inadequate maintenance of the plots.
- **Too many treatments** – trying to compare too many treatments or treatment combinations (for example, several species treated with different fertilizer types applied at different rates, etc.) can lead to data from which making any meaningful conclusions is challenging.

- **Too few trees per treatment** – *it is important to have enough trees (or other study subjects) in each treatment to generate an adequate amount of data from which averages and differences among averages can be calculated with confidence.*
- **An emphasis on being “operational”** – *although the study objective is to generate results that can be applied to operational practices, using an operational approach when conducting the study can result in excess variation. Any variation not attributable to the treatments or subjects being studied makes it difficult to isolate treatment effects and determine the maximum response potential.*

Variation is the Key

Setting up a study of any kind is all about controlling sources of variation. In fact, variation is the basis of most statistical calculations – analyzing variation within and among different groups to determine whether or not the groups differ from one another. For example, if you wish to compare heights for two groups of seedlings (such as groups by species, treatment, or some other factor) and the average height is 22 in (56 cm) for one group and 17 in (43 cm) for the other group, you would then examine the variation to determine if those two groups truly differ in height. If there is very little variation in the data (for example, most height measurements within each group fall within 1-2 in [2-5 cm] of their respective group’s average), then the conclusion would likely be

that the two groups are different. If the data varies quite a bit (for example, some height measurements are much higher and some are much lower than the average) then there is likely a lot of data overlap between the two groups and you cannot conclude that the two groups truly differ in height.

To generate valid and useful data, it is essential to maximize both its accuracy and its precision (figure 3), both of which can be significantly affected by how the study is designed and implemented. Variation created by bias, confounding, or outside influences can generate data that is inaccurate or inconclusive. Ultimately, the only desired source of variation is the variation resulting from the treatments or other factors being studied. Everything else is “noise.”

Since variation plays a fundamental role in the ability to compare different treatments or other factors, proper study design is critical. Understanding and controlling the causes and magnitude of variability are the key to generating data that can be used to make valid conclusions about the treatments or other factors being studied.

Treatments

The treatment is the one factor that is intentionally changed for the sake of the experiment. It is the factor that is expected to create a response. For example, a treatment could be fertilizer rates, fertilizer formulations, growing media components, species, seedling stocktypes, seed lot, planting method, or other treatments. All other factors must stay the same to be able to isolate responses to the treatment in question. So, unless the intent is to compare seed lots, species, planting dates, etc., all of those other factors must be the same throughout the study.

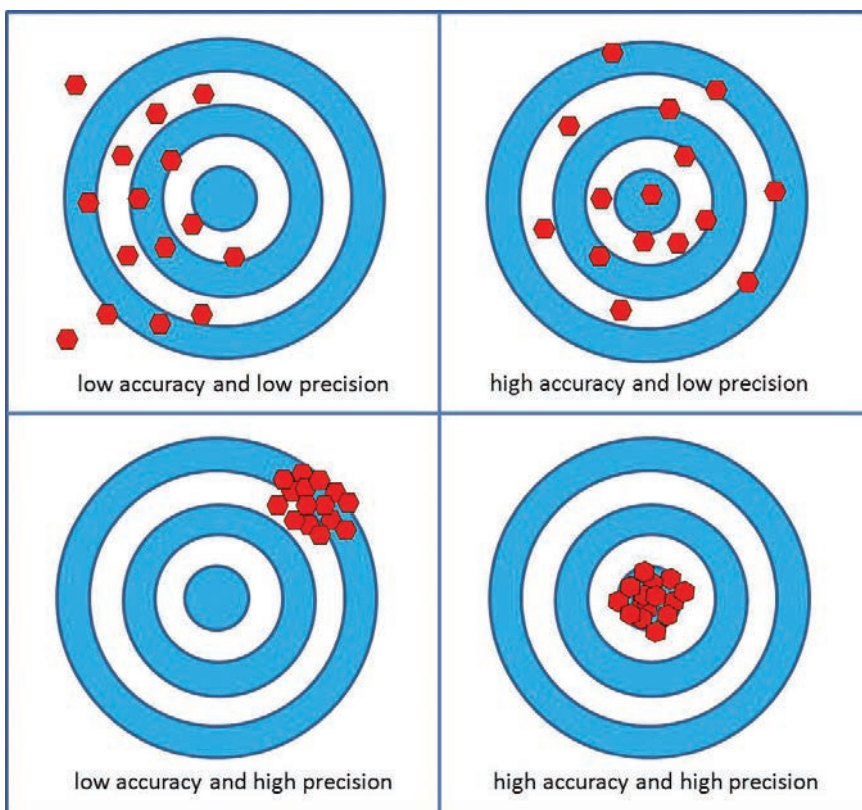


Figure 3. A good study design strives to eliminate bias, confounding, and other sources of variation in order to isolate treatment effects with accuracy and precision.

Control Treatment

Including a control treatment is an essential component of experimentation. The control treatment is the usual method of doing something. It is important to have a control treatment so responses to the modified method can be compared to the usual method.

Some studies may include two control treatments: the “do-nothing control” (where no product is applied to the crop) and the “operational control” (where the usual product or treatment is applied to the crop). Having an operational control is most common with pesticide trials in which new pesticide treatments are compared to the current pesticide in use as well as to a control treatment in which no pesticides are used at all.

Factorial Treatments

Studies can also be designed to evaluate two treatments (factors) at the same time. For example, fertilizer would be factor A and stocktype would be factor B. Factorial study designs allow you to determine if there are interactions between the two factors: is the response to fertilizer the same for every stocktype? For the design to be valid, all combinations of the two factors must be included. For instance, if there are three fertilizer rates (factor A) and three seedling stocktypes (factor B), then there needs to be a total of nine treatment combinations included in the study (3 rates x 3 stocktypes). A control level for each factor must be included as well.

Number of Treatments

While it may be tempting, including more than two factors or more than 10 treatment combinations will not increase the usefulness of a study. Keep it simple—do not include too many treatments and do not go beyond two factors. In fact, increasing the number of treatment comparisons in a study increases the odds of finding a difference when one does not exist. Furthermore, three-way (or more) interactions are very challenging to quantify and interpret. It is better to establish additional studies rather than try to answer too many questions in a single study.

The Three “Rs” of Study Design

Once the objectives have been defined for a study, details about the experimental design need to be established. A good study design does not have to be complicated, but all study designs need to incorporate the “Three Rs” – Randomization, Replication, and Representation. These Three Rs are important tools to control variation and generate valid data that can help answer the questions posed by the study.

Randomization

Randomization is the circumstance in which each experimental unit in the study has the same chance of being assigned to any of the treatments. The experimental unit is the basic unit to which the treatments are applied. This unit must be clearly defined (for example, individual trees, rows of trees, a pallet of seedlings, a field plot, a greenhouse bench, a nursery bed, a greenhouse). Individual trees are good for short-term studies in small areas with relatively uniform conditions. Plots are usually best for forest or nursery studies. The most common plot configurations are row, square, or rectangle plots. Square and rectangle plots are usually better for longer-term studies because they create a very small depiction of how the area would be if it was all treated in the same manner whereas row plots will have a greater influence from adjacent rows.

Randomization prevents bias, which can be defined as any process which tends to produce results or conclusions that differ systematically

from the truth. For instance, if treatments A, B, and C are assigned from left to right to a series of plots, then B is always left of A, and C is always left of B. If there is a gradient in soil or sunlight from left to right, then the trees might respond systematically different due to factors other than the treatment in question.

Following are some other examples of approaches that result in a biased study:

- *“This plot looks weedy; let’s put the vegetation control treatment here.”*
- *“This area is close to the road; let’s install the fertilizer treatments here so we don’t have to carry it up that hill”*
- *“These seedlings are smaller than the others; let’s put them in the plot with the highest irrigation treatment.”*
- *“These seedlings have nice foliage; let’s choose them for foliar sampling.”*

To implement randomization, assign treatments to trees or plots using a random, non-biased method. This can be accomplished by rolling a die, drawing a playing card, using a random-number generator, drawing treatment names/numbers out of a hat, or other methods. To save time and avoid on-the-ground bias, it is best to plan randomization in the office, prior to implementing the study in the nursery or field.

Replication

Replication is the most often neglected, yet most important, component of study design. Replication provides the ability to measure variation whether it is due to the treatments, the study subjects, or the physical conditions on the site. Failure to replicate renders it impossible to make valid comparisons between treatments. Without replication, all you have is a one-time event which may or may not be repeatable. For instance, if a cowboy successfully rides a bucking bull one time, how confident can we be that she or he will do so from now on? Making management decisions based on unreplicated data is just as risky as gambling on the rodeo cowboy who has only ridden the bull once.

Replication is achieved by applying each treatment to more than one experimental unit. As described above, experimental units can be individual trees but are more often field plots, nursery benches, or other units composed of several seedlings. It is important to distinguish that the trees within a plot (or other multi-tree unit) are the sampling units whereas the plot itself is the experimental unit. The most common mistake regarding replication occurs when the sampling units are regarded as replicates when in fact, they are not. This error results in pseudo-replication.

Statistical procedures exist for determining the ideal number of replicates for a given study based on how much variance is expected. Statistical calculations are beyond the scope of this paper, however, and mathematical determinations of study size are not often used for field studies. The most important thing to know is that more replicates are always better than less. Having more replicates (while still keeping the study at a manageable size) increases the study’s ability to detect whether or not there are significant differences among groups. When determining the number of replicates (experimental units) and plot size (number of sampling units), various factors need to be considered such as expected survival, duration of the study, and type of measurements (nondestructive vs. destructive). When individual trees are used as replicates, I recommend a minimum of 25 trees in each treatment (50 or more if possible). When plots are used, I recommend a minimum of 4 plots per treatment, each with a minimum of 10 trees. As stated previously, however, more is better; the study design I have used most often is 5 plots of 25 trees per treatment.

Representation

Common sense tells us to compare apples to apples rather than apples to oranges. This is also a basic tenet of good study design. When designing a study, it's important to be aware of its "scope of inference" - the population and circumstances to which the results can be applied. The study should be conducted such that the results are applicable to the specific trees or situations of interest. For instance, if the objective is to apply the study results to pine trees on high elevation sites, then it would be imprudent to conduct the study with oak trees or on low elevation sites because oak trees and low elevation sites do not represent the situation defined in the study objectives.

To ensure that the study design is adequately representative, select treatments, experimental materials, sites, timing, and situations that best represent the desired scope of inference. By ensuring representation, you can confidently apply the results to specific populations and circumstances.

Incorporating the Three "Rs" into Study Design

There are numerous study designs. For purposes of this paper, however, I will describe the two most common designs used in reforestation and nursery studies.

Completely Randomized Design

The completely randomized design (CRD) is one of the simplest study designs. A representative population of trees (or other study subjects) and site(s) are designated for the study. Within the representative population, trees are randomly selected to be included in the study. These trees are then replicated by individual trees or in plots and randomly assigned to a treatment (figure 4).

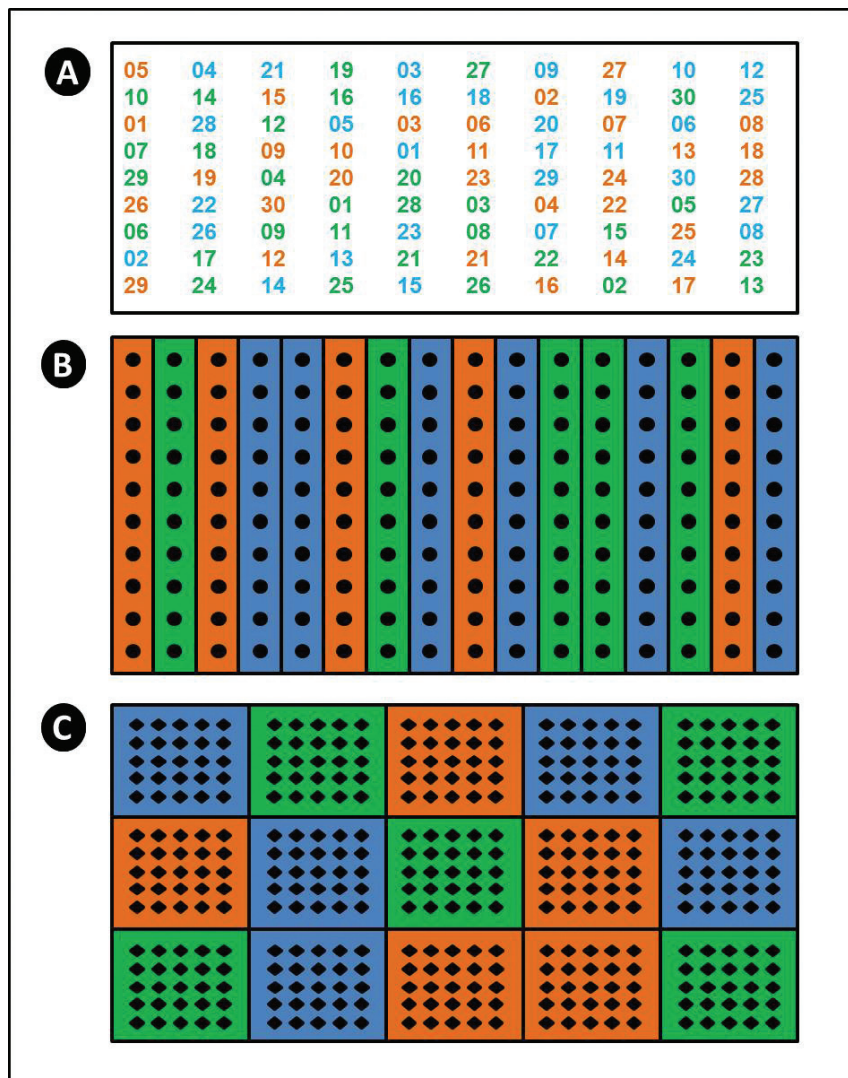


Figure 4. Examples of completely randomized designs to assess three treatments (illustrated here with three shades) using single-tree replicates (A), row plots (B), or square plots (C).

CRD should only be used in situations where conditions on the study site are expected to be homogenous (for example, inside one area of a greenhouse, in a bareroot nursery field, on a flat outplanting site with consistent ground cover, etc.). Although CRD is simple and efficient, it is not often used because researchers are often uncomfortable assuming conditions in their study area are truly uniform.

Randomized Complete Block Design

The randomized complete block (RCB) design is the most common design used in nursery and reforestation studies. This study design can be used under variable conditions (for example, typical outplanting sites, different soil types in a nursery, a series of greenhouses, etc.).

As with the CRD, representative study site(s) are chosen and trees (or other study subjects) are randomly selected from a representative population to be included in the study. These trees are then replicated into treatment plots. One plot of each treatment is then grouped into a block. Trees are randomly assigned to each treatment plot and treatment plots are randomly assigned within each block (figure 5).

Each block in a RCB design is a replicate. For this design to be effective, conditions within each block should be as homogenous as possible but conditions among blocks can vary significantly. Blocks can be located adjacent to one another, spread throughout the site (figure 5), or even established on different sites. Blocking should be based on any condition or gradient that could affect treatment responses (for example, slope, drainage, soil type, aspect, vegetation, etc.).

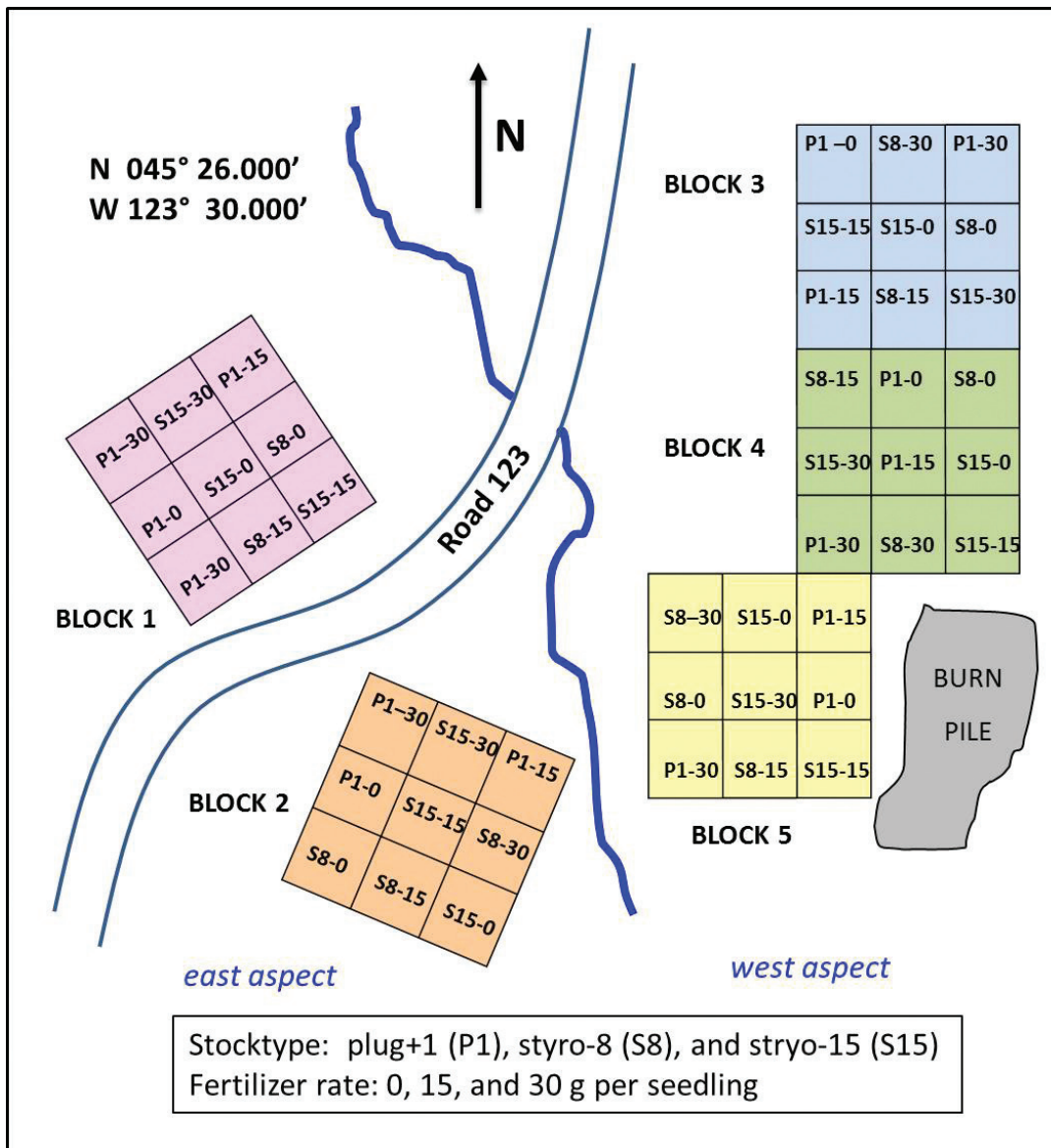


Figure 5. Example of a randomized complete block design with five blocks, each containing nine treatment plots. Note that this example shows 3 x 3 factorial treatments: 3 stocktypes (P1, S15, and S8) and 3 fertilizer rates (0, 15, and 30g [0, 0.5 and 1 oz]). This illustration is also a good example of mapping the site location and layout.

The great advantage of blocking is the ability to perform simple statistical analyses that can isolate the variation due to the treatments in question from the variation due to differences in conditions among blocks (that is, it can separate treatment effects from block effects). The RCB is actually a stronger design than the CRD because the treatments can be compared under a wider range of circumstances; if relative treatment responses are similar in all blocks, even though the rate or magnitude of response may vary due to block conditions, then there can be even greater confidence when making conclusions about treatment effects.

Example #1 Revisited

In Example #1, Cowgirl Jane’s study to test two products in her nursery had a variety of issues (figure 1). First of all, her treatments were confounded with seed lot and with the irrigation pattern in the two nursery beds. Secondly, the seed lots and test location were different than the crop to which she would like to apply the treatments operationally. Thirdly, she did not include a control treatment to enable determination of whether either of the treatments truly is better (or worse) than her existing practices. Lastly, data were collected only on the largest seedlings.

By incorporating the Three Rs into the study design, Cowgirl Jane’s study can be improved greatly. The treatments need to be applied to one representative seed lot in a representative location of the nursery. She can plan ahead to ensure that there will be excess stock available for the study. If she expects seed lots to respond differently to the treatments and wants to include more than one seed lot in the study, then seed lot will need to be a second factor included in the study design (see section describing factorial treatments). She needs to add a control

treatment to the study design and she needs to replicate the treatment plots. If she chooses an area that is relatively uniform (same irrigation pattern, cultural regime, etc. throughout) then she could set up the study in a CRD (figure 6A). Because there can be hidden variation in soil or other factors, however, she may prefer to set up the study in a RCB (figure 6B). Regardless of the study design she uses, the treatments need to be randomly assigned to each plot. These changes to her study design will result in a valid dataset that can isolate the seedling responses to the applied products and determine if they improve crop performance relative to the control. When it is time to collect data, she must randomly select seedlings for measurement from each treatment plot to avoid bias (see later section on Data Collection).

Example #2 Revisited

How can Cowboy Joe incorporate the Three Rs to improve his study design (figure 2)? Because there is a great deal of variation on his site, a good start would be to take steps to reduce variation as much as possible in the study area. He can establish the study plots away from the mature forest to reduce browsing and shading influences. He can also exclude the burn pile from the study area. In addition, he can take measures to control the blackberries. These extra efforts are above and beyond operational practices but are necessary to eliminate excess variation, thereby increasing the data’s accuracy and precision. Cowboy Joe cannot rid the site of all variation (such as soil depth) but by using a RCB design with five replications (blocks) and 20 seedlings in each treatment plot, he can better isolate seedling growth differences due to nursery of origin from growth differences due to site conditions (figure 7). He can also eliminate his own bias about the study outcome by randomly assigning seedlings to plots ahead of time.

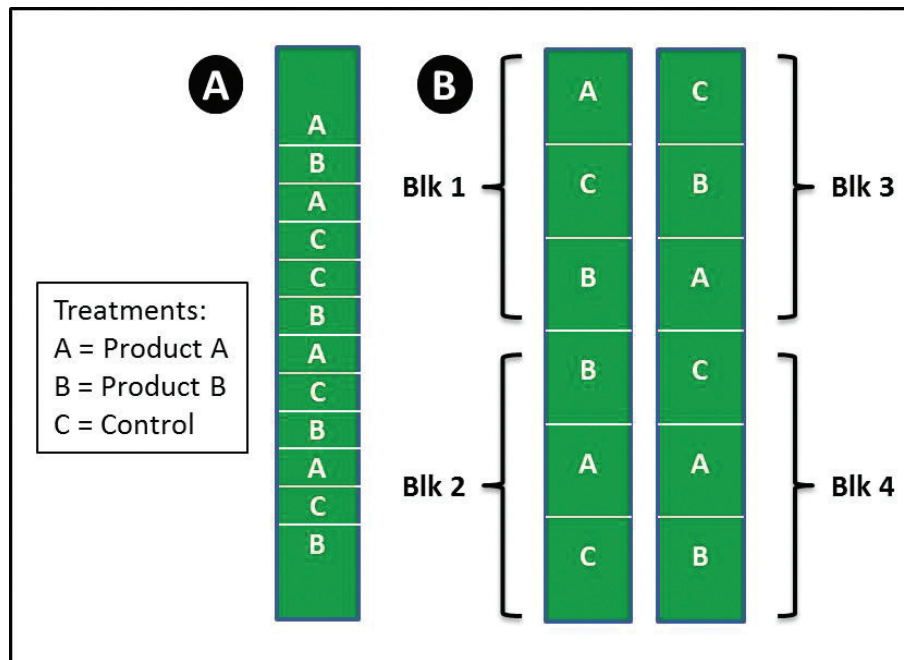


Figure 6. The study design shown in figure 1 can be modified to incorporate representation, randomization, and replication in a completely randomized design (A) or in a randomized complete block design (B) to compare seedling responses to applications of Product A and Product B, thereby eliminating excess variation and confounding. Additionally, a control treatment has been added to determine if either of the treatments is better or worse than the existing method.

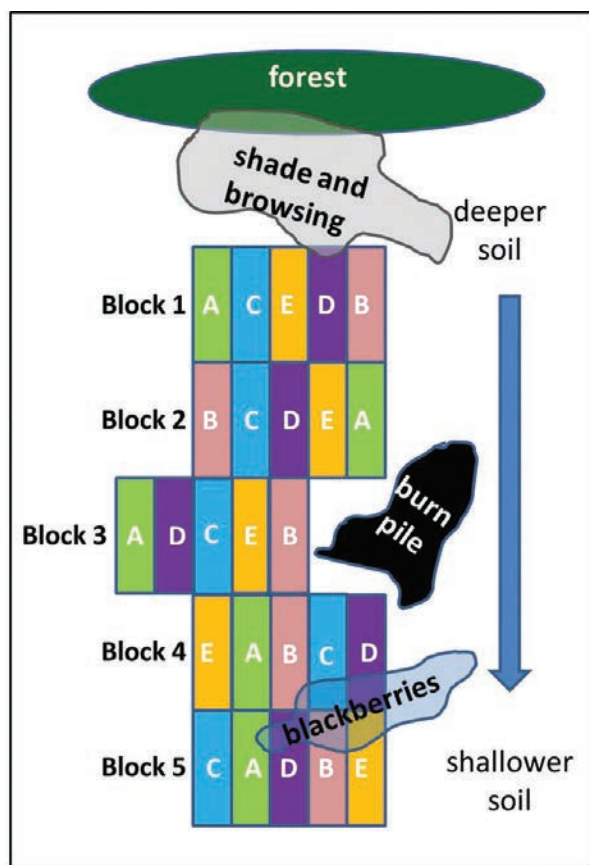


Figure 7. The study shown in figure 2 can be redesigned so that any field performance differences due to nursery of origin (A, B, C, D, and E) can be isolated from variation in site conditions. A randomized complete block design positioned away from known sources of variation or damage along with some vegetation control can improve the quality of the data generated. Note that the revised study design requires the same amount of space and seedlings as the original design.

It's important to note that Cowboy Joe's revised study design requires the same amount of space and seedlings as his original design. There is a misconception that proper study design is costly and time consuming, but this is usually not true. The reality is that poorly designed studies can waste 100% of the time and resources invested, and can lead to additional unnecessary costs if management decisions are predicated on flawed data.

Elements of a Study Plan

Any study should start with a study plan. This document should read like a recipe that anyone can follow from start to finish. The plan needs to be clear, concise, and specific. It does not have to be lengthy but it should contain sufficient detail so the purpose and methods are clearly understood. This is the time to think ahead and plan all aspects of the study. Important elements of a study plan are described in the following sections.

Define the problem and state the objectives

The first step is to describe the issue at hand and the purpose of the study. If the problem cannot be defined, it will be difficult to solve. A

paragraph or two about the problem (history, symptoms, magnitude, consequences, etc.) and the proposed solution will provide the necessary background and justification for the study. From there, the study objective statement can be formed. For example, "The objective of this study is to determine the effect of three fertilizer rates (0, 15, and 30 g [0.0, 0.5, and 1.0 oz]) on first- and second-season growth and survival of Douglas-fir plug+1 seedlings outplanted on a coastal site."

Describe the experimental material and study site

The material selected must be representative of the population in question. For example, "Plug+1 Douglas-fir seedlings (seed lot 123-456, seed zone 071), sown in 2014 at the WeGrow Nursery (Trees, OR), and grown under standard nursery procedures will be used for this study." Likewise, the site should be representative of the environment associated with the problem and objectives. For example, "Seedlings will be outplanted to a site 5 miles NW of Research City, OR at an elevation of 1300 feet. The site was harvested in 2011 and site prepped in 2012."

Describe the treatments

Treatments included in the study should be specific to the problem and objectives. Details about each treatment need to be given. For example, "Four fertilizer treatments will be included in the study: a) unfertilized control, b) 10-25-4 (N-P-K), c) 17-17-17, and d) 15-9-12. Fertilizers are controlled-release (16-month rate) and manufactured by NPK Company (Nutrientville, CA). Fertilizers will be applied once at the time of outplanting, at a rate of 12 g (0.42 oz) per seedling."

Define the experimental design

It is best to use the simplest design that will yield data that can be used to meet the study objectives. Randomization and replication must be outlined. For example, "Seedlings will be outplanted in a completely randomized block design. There will be 6 blocks, each consisting of four treatment plots of 25 seedlings each, for a total of 600 seedlings in the study."

Describe the installation

A good description of study installation specifies dates, labor, equipment, supplies, and any other details associated with establishing the study site. For example, "The study will be planted in February 2014. Color-coded pin flags will mark each planting spot and each seedling will be tagged with block and treatment. Four planters will be needed to install the study and will be monitored for quality. A detailed map of block and plot layout on the site will be prepared."

List the desired data and how it will be collected

It's important to describe the data to be collected on the study including the procedures, timeline, and tools. For example, "Within one week of planting, all seedlings will be measured for initial height and stem diameter. Foliar samples will be collected in July 2014 from a branch in the upper half of 3 randomly selected seedlings in each treatment plot and analyzed for concentrations of N, P, K, Mg, and B. Nutrient analyses will be conducted at Ion Lab, Ltd. (Bunson, ID). At the end of each growing season from 2014 to 2017, all seedlings will be measured for height (groundline to base of terminal bud), stem diameter (1 cm [0.4 in] above groundline), and survival."

Describe how the data will be analyzed

The sources of variation and method of analysis should be determined ahead of time to ensure that the experimental procedures will generate the answers sought. See the Data Analysis section of this paper below for details.

Describe study maintenance and duration

It is important to consider all resources and tasks necessary for the entire study duration. Include necessary annual activities other than data collection. For example, “Competing vegetation will be controlled with herbicide for the first 3 seasons after planting. Plastic mesh tubing and seedling tags will be checked on each measurement date and moved as needed to avoid damage and growth restriction.”

List the expected outcomes

Explain how the study results will be used to address the objective, make management decisions, and determine future research needs. For example, “Results of this study will be used to determine which ponderosa pine stocktype(s) have the greatest growth potential on specific sites in SW Washington. A report of this study will be presented at the 2016 Company Board meeting and an article will be prepared and submitted to *Tree Planters’ Notes* for publication.”

Conducting the Study

A good study design and a detailed study plan can be rendered meaningless if a study is not set up or measured carefully. Use the study plan to guide every step of the study; if anything must be changed, record it in detail. It is important to avoid introducing bias, confounding, or excess variation during study installation or measurement.

Study Installation

Once a study site is selected, the plots should be laid out ahead of time. For an outplanting study, all seedlings should be handled and planted very carefully using experienced planters. As much as possible, the study site should be protected from outside influences that can create more variation and mask potential treatment responses. If browse is anticipated, then the site should be fenced or seedlings protected with mesh tubing. If adjacent treatments have the potential to influence each other, minimize this by installing border rows or buffer strips between treatment plots.

Following is an example of confounding inadvertently created during a study installation: A study plan was developed to compare seedling responses to two different fertilizer treatments and an untreated control using a CRD. The relatively uniform site was laid out ahead of time in a random arrangement of 100 white, blue, and yellow pin flags. In an effort to simplify the planting process, one planter was given a bag of seedlings and a bucket of one fertilizer type to plant at each of the blue pin flags, another planter was given a bag of seedlings and a bucket of the other fertilizer type to plant at each of the yellow pin flags, and the third planter was given a bag of seedlings and no fertilizer to plant at each of the white pin flags. This seemed like a good idea until the forester measured initial height and stem diameter one week later and discovered that seedlings in one of the treatments had a shorter average height than the other two treatments. Since all of the seedlings were from the same seedlot and nursery, and since the sample size was sufficient, this result was unlikely at the onset of the study because treatments could not yet have an influence on seedling size. It turned out that one of the planters tended to plant deeper than the other two planters

resulting in shorter measured heights. To prevent this confounding, the planting could have been done with a single planter or by having each planter plant one-third of the seedlings within each treatment.

Data Collection

As with all other aspects of planning and conducting the study, taking measurements must be done carefully to ensure accuracy and ease of interpretation. It’s important to be consistent when taking measurements (tool used, time of year, and so on). It’s best to measure under ideal conditions if possible; avoid worker fatigue or severe weather conditions to help ensure data quality. Do not introduce any confounding or bias during measurement (some examples: one person measures all of one treatment, or; some treatments are measured earlier than others, or; stem diameter is measured higher up on the stem of trees growing in prickly vegetation).

Initial tree size (or other characteristics of interest) should be measured as soon as possible after the study is installed. This initial data is the benchmark for calculating subsequent changes during the study. Be careful not to damage trees during measurement; broken tops from handling or girdled stems from calipers will result in negative effects on those trees that are not due to the treatment.

If possible, enter data into a spreadsheet on a handheld field device as it is collected. If a handheld device is not available, then carefully enter the data into a computer as soon as possible after it is collected. All data for a single study needs to be in the same spreadsheet so it can be easily analyzed (table 1). Too often, people make multiple spreadsheets for different treatments, different measurements, different dates, and so on. But, data in multiple spreadsheets cannot be imported into statistical software programs and can be unnecessarily confusing.

In addition to measurements on the study subjects, it is valuable to record anything else that may have an influence on the study such as weather events, unusual observations, annual precipitation, etc. It is also recommended to take numerous photos during the study setup and on each measurement date.

Data Analyses

A well-designed study that has been carefully conducted will generate quality data for analyses. Most data for simple field studies as described in this paper are analyzed using Analysis of Variance (ANOVA). Nonetheless, many field and nursery personnel do not have the time or inclination to learn statistical methods nor do they have access to statistical software. Consequently, data sets can sometimes languish or only be analyzed using simple calculations in a spreadsheet. When developing the study plan, it is wise to partner with another person within the agency or company who has a statistical background, with someone outside the company or agency who has access to statistical experience and resources and would like to collaborate on the study, or with someone in academics (professor, student, or extension agent) who can assist with data analyses.

Study Longevity

Accessibility to the site should be available for the duration of the study. A detailed map of the study layout including GPS coordinates, roads, and other major site features is indispensable (figure 5). Also, lasting identification of plot boundaries and individual trees is essential. Pin flags are useful for study layout but can fade over time or be hard to locate once vegetation establishes on the site. Labeled wooden or metal fence stakes can be used to mark the corners or centers of plots. Aluminum tags are useful for tagging individual trees with block, plot, and tree numbers (if placed on the main stem, these tags will need to be moved after a year or two to prevent girdling).

Table 1. A spreadsheet of all data in the study is useful to calculate averages, growth, and ratios and can be imported into software programs to determine if there are statistical differences among treatments. This sample spreadsheet shows data for two plots from a study with two treatment factors (fertilizer x stocktype). The spreadsheet includes the identifying information for each tree (block, fertilizer, stocktype, and tree #) and the height, diameter, and survival data measured just after planting (2/2012) and on two subsequent dates (9/2012, and 9/2013) along with comments (“comm”) for unusual observations (chlor= cholortic; mt = multi-top; dt= dead top). The full data set continues in subsequent rows for all trees in all treatment plots from all blocks.

block	fert	stock type	tree #	ht212 (cm)	dia212 (cm)	comm 212	ht912 (cm)	dia912 (cm)	surv 912	comm 912	ht913 (cm)	dia913 (cm)	comm 913	surv 912
1	con	P1	1	64	9	.	76	11	1	.	107	18	.	1
1	con	P1	2	48	12	.	63	15	1	.	111	29	.	1
1	con	P1	3	56	10	.	66	12	1	.	87	16	.	1
1	con	P1	4	37	7	.	46	7	1	.	70	15	.	1
1	con	P1	5	52	8	.	62	10	1	.	75	17	.	1
1	con	P1	6	57	6	.	.	.	0	dead	.	.	dead	0
1	con	P1	7	51	8	.	59	9	1	.	71	14	.	1
1	con	P1	8	58	9	.	68	9	1	.	82	15	.	1
1	con	P1	9	57	9	.	62	10	1	browse	88	19	.	1
1	con	P1	10	46	7	.	55	7	1	.	67	12	.	1
1	con	P1	11	58	9	.	63	10	1	.	49	18	dt	1
1	con	P1	12	68	11	.	71	12	1	.	83	15	.	1
1	con	P1	13	40	7	.	.	.	0	dead	.	.	dead	0
1	con	P1	14	53	10	.	.	.	0	dead	.	.	dead	0
1	con	P1	15	58	9	.	64	9	1	.	.	.	dead	0
1	con	P1	16	43	6	.	44	7	1	.	43	8	dt	1
1	F1	s15	1	31	5	.	50	10	1	.	66	13	.	1
1	F1	s15	2	23	4	.	43	9	1	.	76	15	.	1
1	F1	s15	3	38	6	.	65	10	1	.	120	21	.	1
1	F1	s15	4	33	5	.	57	10	1	.	93	20	.	1
1	F1	s15	5	33	7	.	52	13	1	.	86	20	.	1
1	F1	s15	6	40	5	.	62	10	1	.	89	17	.	1
1	F1	s15	7	43	7	.	59	10	1	.	73	16	.	1
1	F1	s15	8	43	6	.	75	11	1	.	133	44	.	1
1	F1	s15	9	33	7	.	38	11	1	brown	61	17	.	1
1	F1	s15	10	37	7	.	57	10	1	.	86	17	.	1
1	F1	s15	11	48	7	.	65	11	1	.	80	17	.	1
1	F1	s15	12	35	6	.	37	8	1	chlor	59	14	.	1
1	F1	s15	13	40	5	.	47	10	1	.	88	23	.	1
1	F1	s15	14	37	5	mt	48	6	1	.	54	11	browse	1
1	F1	s15	15	42	6	.	68	10	1	.	74	13	.	1
1	F1	s15	16	41	5	.	53	7	1	.	78	14	.	1
and so on														

References and Additional Reading

- Binkley D. 2008. Three key points in the design of forest experiments. *Forest Ecology and Management* 255: 2022-2023.
- Clewe, AG. 2001. *Practical statistics and experimental design for plant and crop science*. . John Wiley & Sons, Inc. New York, NY. 346 p.
- Cochran WG, Cox GM. 1992. *Experimental Designs*, 2nd edition. John Wiley & Sons, Inc. New York, NY. 640 p.
- Dictionary.com. 2013a. "bias," in *Dictionary.com Unabridged*. Random House, Inc. <http://dictionary.reference.com/browse/bias>. (February 2014).
- Dictionary.com. 2013b. "confound," in *Dictionary.com Unabridged*. Random House, Inc. <http://dictionary.reference.com/browse/confound>. (February 2014).
- Fowler N. 1990. The 10 most common statistical errors. *Bulletin of the Ecological Society of America*. 71: 161-164.
- Pinto JR, Dumroese RK, Davis AS, Landis TD. 2011. Conducting seedling stocktype trials: A new approach to an old question. *Journal of Forestry*. 109: 293-299.
- Rose R. 2000. Cowboy science. *Western Forester*. March/April 2000: 2-3.
- Stafford SG. 1985. A statistics primer for foresters. *Journal of Forestry*. 83: 148-157.
- Steel EA, Kennedy MC, Cunningham PG, Stanovick JS. 2013. Applied statistics in ecology: common pitfalls and simple solutions. *Ecosphere* 4(9): Article 115.