

---

# How to Carry Out an Agricultural Experiment

Edward Berkelaar

---

As you work in agricultural development, there may be times that you find yourself wondering about the answer to a specific question you have. For example, should plants be spaced 30 cm or 60 cm apart to achieve the highest yield? Which one of three tomato cultivars would grow best in a particular area? Would growing a cover crop in the off-season result in higher corn yields? Once you decide on a particular question that you want answered, several steps can (and should) be taken. These steps will make the best use of your time and efforts while giving you the most confidence in your outcome. This article will cover the important steps in planning and carrying out an experiment and then apply these steps to a sample experiment. In some cases we have used big words, but please do not let them turn you off. We have tried to define the words well, and we have highlighted them to make them more obvious.

## **Know your Question!**

The first step is to know exactly what you are asking. The simpler and more specific the question, the better. For example, "Which tomato variety should I recommend in this area?" is a poorly worded question. It is vague and should be narrowed down as much as possible. Perhaps you are in a hot area and already know that you can discount any tomato varieties that were not developed or bred for tolerance to heat. A better question would be "Of the tomato varieties A, B, C, D, and E, which has the highest marketable yield?" The question you ask is closely related to your research hypothesis, which in this case would be: "One of the five cultivars A, B, C, D, and E yields better than the others"; or "Not all of the cultivars have similar marketable yields."

For statistical reasons, it is important to be able to come up with what is called a null hypothesis. This is the opposite of your research hypothesis. In this case, the null hypothesis would be "The tomato varieties A, B, C, D, and E have the same marketable yield." This kind of statement does not seem to make sense, but it is important because use of statistics cannot prove a hypothesis, but it can provide information about a null hypothesis. For example, if the statistical analysis of data suggests that the marketable yields of the different tomato varieties are NOT the same, then you can conclude that the varieties do not all produce the same marketable yield. A similar process can be used for comparing plant spacing, pruning techniques, rates of fertilizer application, etc.

## **Literature Search**

Once you know your question, spend some time looking for information that has already been collected on the subject. Maybe a local research station has done variety trials and the information (or some of it) is already available. Perhaps a variety trial was done years ago or in another location, and you can see how some newly available varieties compare to some others that have been around for a while. You may find guidelines explaining how previous variety trials were done, even if they were for a different crop. Often, the result of a literature search is that you want to modify your question. In the process of doing a literature search, you will become better acquainted with your subject area and end up with a clearer question that you want answered.

### **Plan Your Experiment: Replicate, Randomize, and include a Control**

The next step is to plan your experiment. First of all, what do you want to compare in your experiment? You might want to compare several varieties of a particular species of plant (this is called a 'variety trial'), or you might want to do an experiment that involves treating plants of the same variety in different ways (e.g. you space some 30 cm apart and space others 60 cm apart). In the latter case, each way that you treat the plants is referred to as a treatment.

When planning an experiment, there are three extremely important procedures to carry out: replication of treatments (or varieties), randomization, and having a control as one of your treatments

*Replication*: Replication means that you apply each treatment to several different plants (or rows, or plots) instead of just one. Using two plants, rows, plots, etc. is replication, but is not enough—you should have at least three replicates for each variety or treatment. It is important to replicate within the different treatments because you want your results to be as accurate as possible.

For example, if you want to know if females and males in a population are the same height, the most accurate way to do this is to measure the height of all females and all males, take the average, and then compare them. Clearly, it is not realistic to try to measure the height of all those people. Instead, the population is sampled, and that sample is measured. If you only select one male and one female, you may have chosen a tall woman or a short man, without knowing that these individuals are not 'average.' By replicating (e.g. measuring the height of 8 males and 8 females), you are likely to get a more accurate idea of the average height of a female and a male. It is still possible, though much less likely, that you would choose 8 unusually tall women or unusually short men for your measurements. Replication also provides information about the uniformity of a population. For example, are most women similar in height, or do the heights vary widely?

As another example, assume you have a small field with 10 rows that are each 40 m long, and that you want to know the yield per given length of row of five tomato varieties. One option would be to fill each row with one of the five varieties (Figure 1 a). This way you could plant each cultivar twice, and have two measurements (replicates) per cultivar. Alternatively, since 40 m rows are quite long, you could split them in half (20 m sections), or even quarters (10 m sections) (Figure 1 b and c). This would give you an opportunity to have four, or even better, eight replicates per variety. The only difference would be that instead of yield per 40 m, results would be in yield per 20 m or yield per 10 m. It would involve a little more work because you would need to mark off more sections and make more labels. You would need the

same amount of land and the same number of plants. Statistically, you have increased the power of your experiment enormously. You cannot analyze your experiment using statistics if there are no replicates, e.g. if you plant only one row of each variety and measure the yield of each row. **The more replicates, the better off you are (try to do at least three), although generally, having more than 10 replicates is unnecessary in agricultural experiments.**

For some experiments (e.g. variety trials), it is also important to repeat them in different years to account for differences in growing conditions from one year to the next.

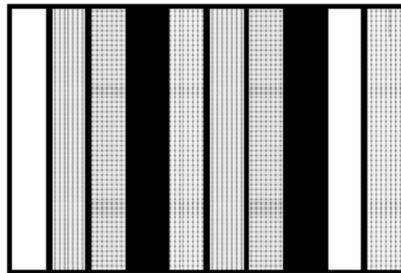


Figure 1a.

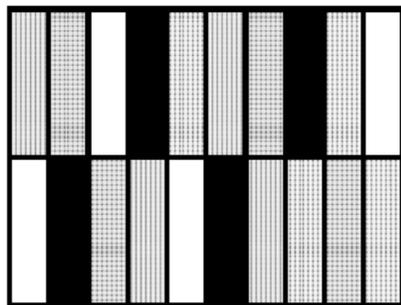


Figure 1b (above) and 1c (below).

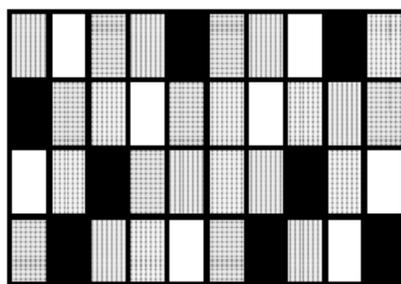


Figure 1. In 1a, ten rows are planted with five varieties (each represented by a different shade), giving two replicates per variety. In 1b, the rows are split in half so that four replicates can be planted per variety. In 1c, rows are split in four so that eight replicates of each variety are planted. 1b and 1c are stronger experimental designs.

*Randomization:* The second important concept is to randomize the location of your various treatments (varieties in this example). This ensures that the different varieties or experimental treatments are planted or distributed randomly, instead of having all of one kind in one place and all of another kind in another place. Randomization is necessary because the growing conditions (e.g. soil environment) in your plot may vary from one area to the next. Maybe a plant variety performed well in your

experiment, not because it was a superior variety but because it was placed where it was more fertile (perhaps fertilizers were not applied evenly or the natural fertility of the soil differed from one area to another). Perhaps one area of the plot was a low point in the field, so that the soil there was wetter. Or maybe one edge of your plot was bordered by a row of trees and received a bit of shade during part of the day. The “magic” of statistical analysis is that it can give you confidence about whether the difference in crop performance you measured was actually due to a difference between treatments or to some other factor.

It is important that conditions be as uniform as possible throughout your entire research plot, but since conditions can never be made exactly the same, it is important to randomly spread differences in your plot among the different treatments.

### **Here is the easiest way to randomize if you want to plant a variety**

**trial.** First, mark out as many planting beds as you need (the number of varieties that you are testing multiplied by the number of replicates). Next, write the name of each variety on a small piece of paper. For each variety, you will need as many slips of paper as there are replicates. Next, put the slips of paper in a bag. Then go to your first planting bed and remove a paper—that is the variety you will plant first. Continue doing this until all the varieties are planted.

*Use a control:* A control is the variety or treatment to which others are compared. It is important to include a control as one of your treatments, and sometimes it is useful to include more than one. Imagine an experiment in which a new growing technique is tested and results in an excellent crop yield. Including the old growing technique as a control allows you to determine if the high crop yield was due to the change in growing technique or to another factor such as an optimal growing season. If you want to do a variety trial, it is always good to include at least one commonly grown local variety. Since controls are exposed to the same conditions (both good and bad) as your other treatments, they serve as an excellent point of comparison. Controls should be replicated and otherwise treated the same as your other treatments. A control is essential; it would not be acceptable to simply compare your results to data from yield of a previous year, or to compare your results to published data. (It is okay to compare data to published data, but not to do that instead of having a control.)

### **Record Observations & Data**

**A written report** of your method and of the final results is important if you want to share this information with others—or even remember it yourself in future years. Others may try your technique, and it may not work. In such a case they will be very interested to know why not. What type of soil do you have? What were your weather conditions like? What time of year did you do your experiment and how long did it last? Did you fertilize your soil and, if so, when? With what kind of fertilizer, and how much of it was used? Did your plants suffer from any type of disease or from any pests? Information like this might explain why an experiment led to different results when it was done at a different time or in a different location. For example, if two tomato variety trials were done, it would be informative (but also a bit confusing) to know that in the first trial, Variety A did best and in the second trial, Variety D did best. It would be helpful to know that during the first variety trial, weather was ‘cool and damp’ while in the second variety trial conditions were ‘hot and dry.’

At the end of your experiment, **record your data.** The way you measure yield should be chosen carefully to ensure that it answers the question you are asking. Make sure you treat all of the plants in the experiment the same. Harvest everything at the same time if possible, or if this is not possible, try to harvest 25% of each treatment rather than everything from one treatment one day and everything from a second treatment the second day. If more than one person is harvesting, explain to everyone the standard used to decide whether fruit should be harvested, discarded,

or left on the plants for future harvests. With more than one worker, it is also advisable to switch halfway through harvesting a treatment, so that one person doesn't harvest treatments A and B only while the second person harvests C and D only. This can be another source of error when you are analyzing results; perhaps one person is a sloppy harvester, or has a different technique than the other.

### **Summarizing your Data: Statistics**

Statistics is a way to summarize data. It is important to understand what statistics can and cannot do. Statistics relies on probabilities. It can allow you to know if the averages of two columns of numbers (treatment 1 and treatment 2, or variety 1 and variety 2) are different from one another. Statistics will give the answer to that question along with a probability. In agricultural experiments, that probability is set at 0.05 or 0.01, meaning that although you might conclude that the averages are different, there is a 5% or 1% chance that your conclusion will be wrong. This is a fairly small chance. In contrast, you would not have confidence in a conclusion that had a 25% chance of being wrong (a probability of 0.25).

For example, if you have two averages, 9.2 and 12.6, are they statistically the same or different? The answer to this question depends on two things; the difference between the two numbers (3.4 in this example), and the variability in the numbers the average came from. If 9.2 were the average of 8.2, 9.0, 9.7, and 9.9, while 12.6 was the average of 10.8, 11.7, 12.9, and 15.0 (i.e. in each case, the numbers were similar to the average), then we might conclude that the averages were not the same. On the other hand, if 9.2 were the average of 4.7, 5.8, 12.3, and 14.0, and 12.6 was the average of 3.9, 9.1, 16.5, and 20.9 (i.e. the numbers that make up each average vary widely), then we are faced with a different situation, and we could not conclude that 9.2 and 12.6 were statistically different from one another.

### **Write a Report**

Once the data have been collected and analyzed and conclusions have been drawn, it is important to **write a brief report**. The report should contain several sections. In the Introduction, it is important to include the question you asked, why it was important, and any additional relevant information that you discovered while you were doing your literature search. The second section is called the *Materials and Methods* section, and should describe exactly how you carried out the experiment (the materials and methods you used to actually do the experiment). This section should be written in enough detail that someone could repeat your experiment using your description. The final section of the report is called the *Results and Discussion* section, and contains the data you collected along with conclusions you drew. Results from statistical analysis are typically included here, along with any ideas you might have regarding why the results came out the way they did. At the end of the report it is important to list any publications you referred to, so that others reading your report may also find and refer to them.

### **Example experiment**

The data below are from an experiment that was actually done at ECHO, but we have simplified it by reporting results from only three varieties here.

Question: Which of three different tropical pumpkin (*Curcubita moschata*) varieties ('La Primera', 'Butternut', and 'Acorn') has the highest yield?

Research Hypothesis: One of the three tropical pumpkin varieties ('La Primera', 'Butternut', and 'Acorn') has a higher yield than the others.

Null Hypothesis: The yields of the three tropical pumpkin varieties ('La Primera', 'Butternut', and 'Acorn') are the same.

Experimental Design:

Number of Plants: 54 (18 of each variety); 3 varieties replicated 3 times (each replicate was a bed of 6 plants)

Treatments: 3 different tropical pumpkin varieties; 'La Primera', 'Butternut', and Acorn. 'La Primera' was the control in this experiment, because it is a variety that is grown commercially in Florida.

Randomization: The experimental design was a completely randomized design (CRD). Other experimental designs exist and are useful in certain circumstances. The CRD is the simplest, most straightforward design. Following is a description of the easiest way to randomize this variety trial. First, make nine planting beds, each of sufficient size to contain six plants. Second, get nine slips of paper and write 'La Primera' on three of them, 'Butternut' on the next three, and 'Acorn' on the last three. Mix the papers in a hat or bowl and draw them out one by one. The order in which the papers are drawn is the order in which the different varieties should be planted

	Yield of six plants in kg		
	'La Primera'	'Butternut'	'Acorn'
Replicate 1	21.1	4.42	15.2
Replicate 2	17.9	2.95	5.78
Replicate 3	21.2	12.8	13.1
Average	20.1	6.72	11.4
Standard error	1.08	3.07	2.86

Table 1. Yield data for tropical pumpkin variety trial.

*Data and Analysis:*

Data are shown in Table 1, below. Pumpkin yields (of six plants) are shown for each variety and each replicate. Yields of six plants averaged over all three replicates are also shown, and so is the standard error for each variety.

Standard errors are a measurement of variability within a variety. For example, the three yields of 'La Primera' are quite similar, and its standard error is small, while the yields of 'Butternut' are not as similar, and its standard error is higher. The smaller the standard error, the more uniform the data.

The next step is to determine if, statistically speaking, the average yields of the different treatments are significantly different from one another. In our example, are the differences between the averages 20.1, 6.72 and 11.4 due to the fact that the different varieties actually yielded different amounts of pumpkins (i.e. are they 'significantly different'), or were the differences due to chance? A statistical analysis will indicate which explanation is most likely. An explanation of statistical analysis is beyond the scope of this article, but the outcome of a sample analysis is presented below. We are putting together additional information that will be helpful if you plan on doing statistical analysis of your data. We can mail or e-mail the information to you. We will also post it on our web site when it is ready.

A statistical analysis was done on the above data to test our null hypothesis that the yields of the three tropical pumpkin varieties ('La Primera', 'Butternut' and 'Acorn') are the same. The analysis can test whether or not the null hypothesis is true. A low probability or p-value ( $p < 0.05$ ) means that the null hypothesis is not true and that at least one of the varieties had a different yield than the others.

In this case, we compared each pair of varieties, resulting in the following p-values:

'La Primera' vs 'Butternut'  $p = 0.015$

'La Primera' vs 'Acorn'  $p = 0.046$

'Butternut' vs 'Acorn'  $p = 0.33$

Typically, the cutoff p-value is 0.05. This means that when the probability is higher than  $p = 0.05$  (5%, or 1 time in 20), we do not have the confidence to say that the averages are different. In this case, we can conclude the following: 1) The yield of 'La Primera' was different than that of 'Butternut' ( $p = 0.015$ ) and 'Acorn' ( $p = 0.046$ ); and 2) the yield of 'Butternut' and 'Acorn' did not differ. In the case of conclusion 1, there is less than a 1 in 20 chance we are wrong. Because the first comparison has a lower p-value (almost 1 in 100), we are more confident in it than the second comparison. Still, there is a very small chance that we are wrong—these kinds of analysis can never allow us to be absolutely sure. [Ed: scientists are sometimes the objects of jokes because they are so precise and hesitant to say much with certainty. Someone described a scientist as a person who, if asked what color a certain house is, would say, "The side facing me is yellow."] We cannot say that the yield of 'Butternut' and 'Acorn' differed because there is a very high probability (33%, or 1 in 3) that these averages are different due to chance. This is probably due to the relatively high amount of variability in the yield data of 'Butternut' and 'Acorn'.

*Conclusions:* We can reject our null hypothesis that "The yields of the three tropical pumpkin varieties ('La Primera', 'Butternut', and 'Acorn') are the same." Our statistical analysis showed that the yield of 'La Primera' was probably higher than the yield of the other two varieties. There was no significant difference in the yield of 'Butternut' and 'Acorn', despite the rather large numerical difference measured; there was a high probability (1 in 3) that the difference in these yields was due to chance.

In this case, the probability that the averages are all similar to one another is 0.0004 ( $p = 0.0004$ ), or 4 in 10,000. We can confidently say that at least one of the treatments was different than the others