

## Research Methods: Experiment Design Lecture

**The Great American Lab Experience, Reading a Lab Manual.** Let's consider the traditional approach to performing an experiment in an academic setting. Preparation consists of purchasing a lab manual, a notebook and some safety equipment (goggles, lab coat), followed by a reading of the lab experiment to be performed. Often, proof that you have read the material (and maybe understand the lab to be performed) consists of writing an "Objective" and "Procedure" in the lab notebook. You then enter the lab, are admonished to wear your goggles, and begin the lab by interpreting step 1 from the written description of the lab, completing it, and proceeding to step 2, and so on, until you have finished step n, where n is the last step in the lab. You may not have noticed, but everything you needed to complete the lab was at the ready. A teacher or T.A. was available to answer the question, "where is the ?" by either pointing to it, telling you to visit the stockroom, or asking you to reread the manual. You also may not notice that, for the most part, everything you are asked to do, works. That is, everything you are asked to do has an expected outcome that will serve you well when it is time to write-up the lab. If you do not obtain the expected outcome, something is "wrong" and you have to redo those steps of the procedure. You may also notice that it is absolutely essential that the experiment be completed within a specified two or three hour time period. Push too long past this time and your teacher or TA will be sure to let you know it.

What you probably do not consider while completing this traditional lab experience is that someone has gone to great lengths to create a lab manual that includes experiments that allow you to successfully complete a desired objective within a specific time period using available equipment and supplies. It is this individual who has to take an infinite number of factors and restrict them in order to create a narrowly constructed lab experience that will be successful. What this person is doing falls in the general category of Experiment Design.

**Research Methods Lab Experience, Write Your Own Lab Manual.** One of the important purposes of this course is to allow you to experience some of the important features of **experiment design**. Of course, teaching experiment design to you is a lot like teaching someone to drive a car. You can sit and explain and explain and explain, but until you drive the car, you aren't going to have much of an appreciation for the experience. Another problem with learning about experiment design is that it just seems so logical, so obvious, that it can't be that hard. In fact, it is an incredibly logical procedure--the problem is that it is being applied to the world in which you live, and in case you haven't noticed, trying to control anything in this world is extremely difficult.

So as you listen to this lecture material, looking at the big picture, **you are learning how to write a lab manual.**

**Experiment Design Questions.** This section of notes is devoted to several questions that arise in the process of designing an experiment. The process of experiment design begins once you have decided upon a hypothesis. It is at this point that you ask yourself several fundamental questions about how to create an experiment.

- How well do you understand the theoretical foundation for the experiment to be performed? Are you aware of mathematical models that are describing the phenomenon of interest?
- What is the response that you will measure? What device will be used to measure the response?
- What factors contribute to the response? Of these factors, how significant is each in contributing to the response?
- What is the relationship between the factors to be evaluated? What is the certainty that factors are independent of one another?
- For the factors to be held constant, what is the reliability with which the factors can be controlled? What experiments should be conducted to evaluate the control data? What sequence will be used for collecting the data? Is there merit in collecting the data as a random function of the variables or is experiment time and reproducibility an issue?
- What is the precision with which the data is collected? How much data is to be acquired to achieve the desired statistical validity?
- How much of the response surface is to be mapped? Specifically, for the factors to be evaluated, what is the range of data that can be collected? What is the resolution with which the data is collected?
- Is an optimization to find a response surface maximum or minimum required? What optimization procedure is to be used?

**Experiment Design Answers.** A qualitative discussion of each of these questions is now presented

- How well do you understand the theoretical foundation for the experiment to be performed? Are you aware of mathematical models that are describing the phenomenon of interest?

Collecting data is easy to do--just read numbers off of a measuring device at various times and write them in your notebook. I once had a graduate student collect hundreds of data points on a spectrometer with the machine in the "Off" position. He was a little confused about why the numbers were always zero, but it didn't stop him from collecting away. You had to admire his industry; he went on to get a Ph.D. in just three years.

Collecting data in the midst of ignorance is about as useful as collecting data with the machine turned off. To shed light on the relevance of the data, and to provide you with a

clear idea of what kind of data you should collect, it is important to know as much as possible about the experiment. Often this means engaging the literature to develop a foundation in the subject matter and to understand what the current knowledge is. For our purposes it is often enough just to provide a foundation. As your expertise grows in an area, you can begin to compare your results with current knowledge and perhaps even add to it.

It is also desirable to begin to develop a mathematical foundation for the data you are collecting. Often times it comes in the form of a simple model that offers a good approximation for the data you will collect. You will recall that the ideal gas law was derived as a simple model to relate pressure, volume and temperature to the number of gas molecules. More sophisticated models corrected for additional factors as the model became more accurate.

Is it really so terrible to start collecting data with little foundation? For many of you, little outside preparation accompanied your first efforts in the Research Methods inquiry labs. In effect, you just came into the lab and played around with the experimental apparatus. You probably noticed, though, that it was as a result of this play that you began to formulate your own ideas for new experiments, began to consider which factors were important, began to get a feel for the system response to each factor. From this, perhaps you even developed hypotheses that would prompt additional experimentation.

Note here the interplay between inductive and deductive reasoning. Actually there is considerable debate among scientists about how much time should be devoted to developing a good foundation in a subject area, and how much time should be spent just playing around in the lab. Some would argue that if you spend all your time reading the literature, it narrows your focus and you increasingly adopt lines of discovery that are similar to those that you've read about. Creativity is squashed.

There is no simple answer that can be given to this issue. Obviously you are better off if you start reading the directions on how to use an instrument than if you just started tinkering with it. But on the other hand, hours and hours of library work are not necessarily going to provoke you in the same way that hours and hours of collecting data will.

- What is the response that you will measure? What device will be used to measure the response?

At the heart of data collection, you have a system that you wish to investigate and there is some information about the system that you wish to generate. This information is called the system response. You may think of the response in terms of an x-y plot--it is what is plotted on the y-axis. You have probably thought of a system response in terms of independent and dependent variables--the response is the dependent variable. You may have thought of the response in terms of cause and effect--it is the effect.

As you plot the response as a result of changes in the system, you are creating a response surface. The response surface is a map of the interesting information about the system. It is your job in experiment design to create this response map in an efficient and unambiguous manner. The rest of the discussion about experiment design considers the issues in how this is accomplished.

When you have a choice, your response should be capable of achieving as large an amplitude as possible. The more the amplitude of the response (the greater the signal to noise), the more information that you can generate. Often this is a function of the equipment that is used. For example, you might use your ears to hear a sound, but could use an amplifier to increase the amplitude of the sound and thereby improve your ability to discern differences in the sound. You might use your eye to measure a length, but use a microscope to amplify your ability to discern differences in length.

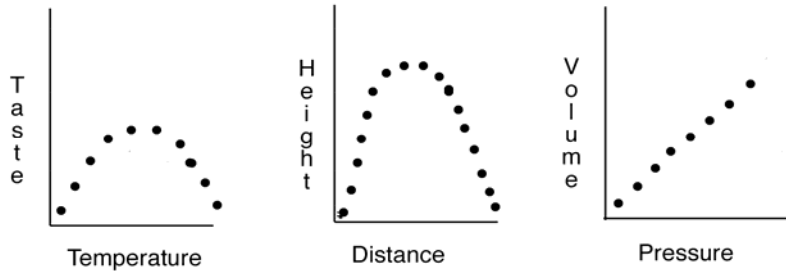
It is also important to be able to calibrate your response so that you can provide a quantitative measure of it when graphing data. You are familiar with various kinds of measuring devices, graduated cylinders for volume, thermometers for temperature, and rulers for length. Each of these measuring devices provides a quantitative measure of the response. Obviously the better the quality of the measuring device, in terms of range and resolution, the better your ability to measure response.

Often you are called upon to measure responses that are subjective in nature and not easily amenable to measurement. The most famous example of this is the hardness scale for rocks, for which a ranking from 1 to 10 describes the increasing hardness of an object. You may certainly create your own scale for a subjective measurement; it is done all the time in the behavioral sciences and sociology. For example, teacher evaluations are scaled from 1 to 5 with 1 being very poor and 5 being excellent. When it is not possible to use a measuring device, a ranking scale provides the ability to graph your data and more readily extract quantitative about your response surface.

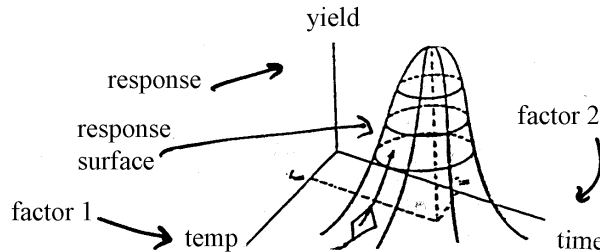
- What factors contribute to the response? Of these factors, how significant is each in contributing to the response?

Perhaps the most important consideration in evaluating the response of a system is the identification of factors that contribute to the response. Factors are also referred to as independent variables. They are the values typically plotted on the x-axis. They are independent variables that you hope to control in a systematic way as you measure changes in the response surface.

As we will see, in the ideal experiment we would like to identify a single factor that determines the response in a system. This allows us to create a simple x,y plot from a collection of (x,y) data points.



It is also not uncommon to imagine plotting data as a function of two factors. This three dimensional plot creates a familiar response surface.



Note that if more than three independent factors are varied, it is not possible to represent the data in a single plot. However as you learning in multi-variate mathematics courses, it is possible to perform mathematical analysis in  $n$  dimensions for  $n-1$  independent factors. Computer programs work well with  $n$ -dimensional arrays and make it easy to perform mathematical computations on the data.

The description of data provided above is an ideal. Although in practice we want to be able to perform experiments in which a single factor is changed, there are many complications to this ideal. For one, the numbers of factors that can possibly contribute to response is often enormous and it becomes necessary to hold all the factors except one constant when creating a response surface. Another complication is that factors are not always independent of one another. Often when we vary a factor to generate a system response we are creating a change in another factor that we are trying to hold constant. Let's look at these to issues in more detail.

Although it might not be apparent, there are an infinite number of factors that can perturb your system. Many of them are ignored because your response is not sensitive enough to measure them. For example, you may want to measure the contributions to sound at a rock concert and are considering the contributions of the instruments and the audience yelling, but ignore the contribution from the car door slamming on the street. However if you had a more sensitive detector of sound you might have been able to account for the car door.

To give you an idea of how complicated factors are, think about something as simple as making a delicious oatmeal raisin cookie. I know for a fact that the finest oatmeal cookies I ever baked were the night before the Raiders kicked butt in the 1981 Super Bowl. Everyone at the Super Bowl party commented on how chewy and delicious the cookies were. For the past 20 years I have been trying to reproduce that experience (no, not the Raiders winning), and I cannot. I have tried everything. All kinds of different flours. All kinds of different combinations of oil and shortening. All kinds of different

ways of mixing the ingredients. More egg, less egg white, more mixing, less blending. And I have never since made the perfect cookie, not even close.

The problem is that there were too many factors involved in making the cookies, and I was even following a recipe! Imagine how complicated the search for the perfect cookie would be if I wasn't following a set of instructions that narrowed my investigation.

It is interesting to think of factors in terms of the kind of experiment being performed. Different branches of science afford varying levels of ease in identifying and controlling factors. The behavioral and social sciences must deal with an incomprehensible number of factors, and as such, rely upon sophisticated statistics such as factor analysis, to interpret the factors that produce a response (they still guess most of the time). The study of living organisms in biology is also subject to enormous numbers of factors, making experimental design extremely difficult. Chemical systems are increasingly controlled in the laboratory, although, the baking of cookies falls into the category of chemistry and you see how complicated it is. This of course, leads us to physics. Many of you are working with physics instrumentation and will find that it has been devised to try to reduce or even isolate a single factor contributing to a response. With a little consideration, though, you can find that a great number of factors are being controlled to create the simple measurement that you perform.

- What is the relationship between the factors to be evaluated? What is the certainty that factors are independent of one another?

The idea of factors is made more complicated by the fact that one variable may affect another variable. In this case one factor is dependent on another. For example, suppose you are trying to grow the ideal lawn. Two variables that affect the grass are water and weeds. Both water and weeds are independent variables that affect the growth of the grass. However water also affects the growth of weeds which means the weeds which affect the grass are themselves dependent upon the water. It is no easy task to separate out the contribution that each factor makes to the growth of grass. A process called factorial design in which each of these factors is systematically varied and the response is measured can be used to distinguish relative contributions. You should be happy I will not discuss factorial design in this course. Of course, ideally you would like all variables to be independent of one other and often strive to create such conditions in the laboratory.

- For the factors to be held constant, what is the reliability with which the factors can be controlled? What experiments should be conducted to evaluate the control data? What sequence will be used for collecting the data? Is there merit in collecting the data as a random function of the variables or is experiment time and reproducibility an issue?

An important feature in experiment design is the need to hold constant all factors except the one being varied to create a response. The expectation that all factors but one are being held constant is often assumed, but rarely evaluated. For example, if data collection takes place over a lengthy period of time, the possibility of a drift in factors you want to control is certain to occur. A common example of this is temperature drift. Imagine you are trying to measure the pressure-volume relationship of balloons. You assume that the room temperature is constant but over time, as the day becomes warmer, or as your body temperature increases the room temperature, there is a continuous increase in temperature that impact the results. For this reason it is common to collect data in a random fashion so that systematic drifts in control variables is averaged out.

This idea of drift in the controlled variables introduces the more general consideration of the reliability of the control data. Most often control experiments are not conducted, usually with the false optimism that the control data will not vary appreciably. However the measurement of the reproducibility of control data provides a statistical foundation when evaluating small changes in response when a factor is varied. After all, not every measurement yields a huge response and without careful control measurements, useful data is not validated. Consider the example of the radioactive  $^{14}\text{C}$  in  $\text{CO}_2$  taken up by a plant. In that experiment the amount of  $\text{CO}_2$  uptake was not enormous--it didn't blow the Geiger counter off scale.

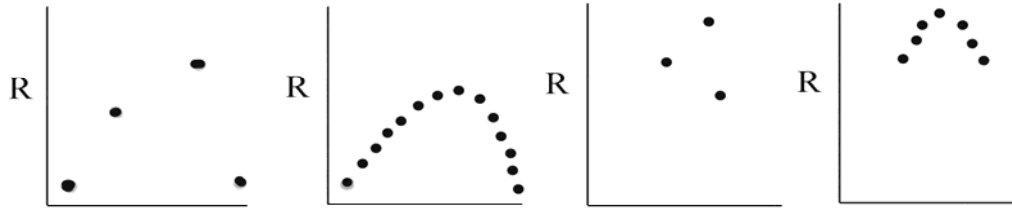
So how might one determine with reliability determine that there was  $\text{CO}_2$  uptake? In the experiment design, several background measurements were made to establish a control population. Then several measurements of radiolabeled  $\text{CO}_2$  uptake were made. Even though the measured signal was small, by collecting an adequate number of control and experiment measurements, it was possible to perform a t-test to demonstrate that the two sets were systematically different.

- What is the precision with which the data is collected? How much data is to be acquired to achieve the desired statistical validity?

The example presented above pointed to the advantage of collecting multiple sets of data. As the expression below indicates, the more data that is collected, the better the random distribution of data is known. This increased reliability makes it possible to distinguish between populations or to maximize the reliability when compared to a known standard. There are practical limits to the amount of data collected--experiment time, the cost of the equipment, and the introduction of systematic error. To the extent that these constraints can be overcome, there are also diminishing returns in the collection of large data sets to establish a statistically significant data set. For example, an examination of  $t$  for sampled data indicates that beyond about 25 to 30 reproducible data points, there is little statistical gain in collecting more data.

- How much of the response surface is to be mapped? Specifically, for the factors to be evaluated, what is the range of data that can be collected? What is the resolution with which the data is collected?

Once the factor to be varied has been identified, it is of interest to consider the range and resolution of data on the response surface to be collected. Consider the following graphs of x,y data points.

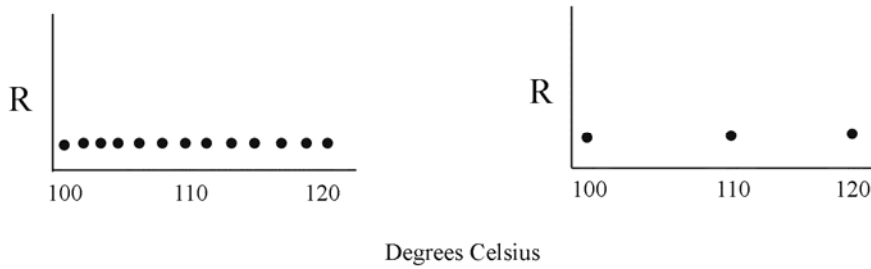


In the first case a wide range of data is acquired with poor resolution. In the second case a wide set of data is acquired with good resolution. In the third case a narrow range is collected with poor resolution. Finally, a narrow range of data is collected with good resolution. Anytime you acquire data, you must choose the range and resolution of the data. What considerations are there in these choices?

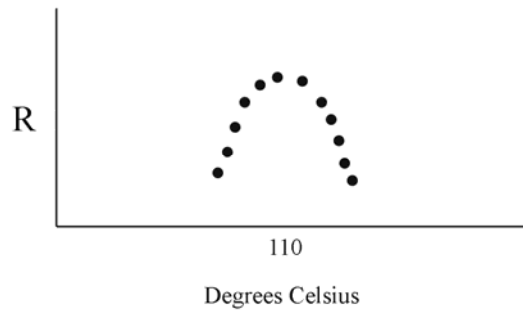
First consider practical issues. The choice of range is often determined by the instrument used to make the measurement. A balance might measure from 0 to 400 grams. An oven might measure from 150° to 450°. The resolution achieved might also be dependent upon the measuring device. Perhaps your meter stick only measures to the nearest 1 mm or the voltage is only varied with a single turn pot. These instrument limitations often determine whether you will be able to perform an experiment to acquire the information of interest to you on a response surface. If there are limitations, the purchase of a typically more expensive device is often the only option.

If one assumes that the appropriate range and resolution are accessible with available equipment, the experimenter must still consider how much data is to be collected. There is little to be gained from having collected 100 data points when just as much information is obtained from 10 data points. As a rule of thumb it is appropriate to collect the largest amount of data in regions where the greatest amount of information is being generated. For example, if one is varying the temperature over a 100° and there has been no change in the system response, is it necessary to collect a data point every degree, or would it be fine to collect a point every 10°?

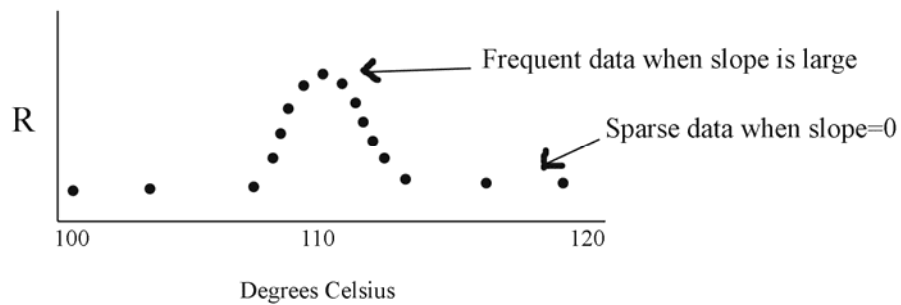




On the other hand, if the slope of the response is varying quickly, for example near a peak in the response surface, it is generally better to collect data with higher resolution so that the changes are better differentiated and the ability to identify the peak maximum is improved.



The ability to collect data in this fashion, with large steps when the response surface affords little information, and small steps when the response surface is interesting is a trait of all good scientists. It is not uncommon to perform an initial survey of the response surface, identify regions of interest, and then fill in data in regions of interest



- Is an optimization to find a response surface maximum or minimum required? What optimization procedure is to be used?

Let's devote an entire lecture to this subject.