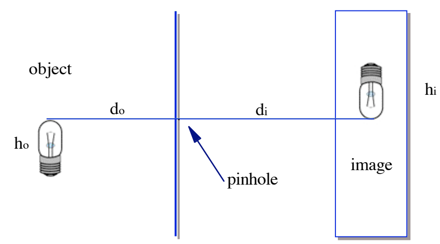
Generic Experimental Design

While scientists will generally agree that there is no "scientific method" per se, there is some basic agreement about procedures that can be followed for experimental design. For instance, consider the following steps:

1. Identify the problem to be solved.
2. Identify the system being studied.
3. Identify and distinguish system variables.
4. Identify the general procedure to be followed.
5. Identify the model if possible.
6. Choose the range for the variables.
7. Collect and interpret data.
8. Consider the overall precision of the experiment.

These steps, if nothing more, make it clear that a number of important factors need to be considered for generic experimental design. Let's examine each of these factors in turn using a generic example.



Step 1. Identify the problem to be solved. A physicist wants to experimentally determine the relationships associated with image formation in a pinhole camera (see the image above). The physicist seeks to find all the relationships between the height of an object, h_o, the height of its image, h_i, the distance of the object, d_o, and the distance of the image, d_i, as measured from the pinhole.

Step 2. Identify the system being studied. Just how the physicists solves the problem will depend upon the system being studied. Clearly, this problem deals with pinhole projection, so the system will consist of a light bulb (the object), a pinhole in a screen, and a screen where the image will form.

Step 3.Identify and distinguish system variables. Four of the system variables have already been clearly identified: h_i, h_o, d_i, and d_o. Are these the only system variables? No. There are other things that might vary in this experiment such as the size of the pinhole, the shape of the pinhole, the number of pinholes, the materials out of which the system is made, the kind of light bulb used, the nature of the material in which the pinhole is made and so on. Some variables are pertinent (e.g., h_i, h_o, d_i, and d_o), some are extraneous (e.g., probably the materials of which the system is made, the kind of light bulb used, the nature of the material in which the pinhole is made). There are other variables that are pertinent and must be controlled (e.g., the shape, size, and number of the pinholes). Other variables will be allowed to change and will be identified as either dependent or independent variables.

Step 4. Identify the general procedure to be followed. Scientists conduct controlled experiments. In controlled experiments there will be only one independent and one dependent variable at any one time. Other variables are held constant and are considered during that phase of experimentation to be state variables. The reason we have only one independent and one dependent variable at a time is so that we can determine the unique relationship between these two variables. If two or more variables are allowed to change independent, there is no way of telling how much affect each of the independent variables has on the dependent variable. Clearly, in our experiment we need to allow four variables to change during different phases of the experiment, but this must be done in such a way that there is only one independent and one dependent variable at a time. The other pertinent variables will be held constant. To begin the experimental study, the physicist decides to see what how varying d_i affects h_i. During this phase of the experiment d_o and h_o are held constant. During additional phases of the experiment there will be other combinations of h_i, h_o, d_i, and d_o always with one independent, one dependent and two controlled variables.

Step 5. Identify the model if possible. Some times a theoretical analysis of the system will point to expected outcomes. This analysis can help interpret the data. In our hypothetical experiment, the physicist relies upon knowledge of the straight-line propagation of light and geometry to predict the outcome of the experiment. The real problem then is to determine whether or not experimental evidence supports the theoretical model. Sometimes it is not possible to directly predict the relationship being studied using theoretical approach, and such pragmatic approaches as dimensional analysis may be used to at least get a general understanding of what to expect.

Step 6. Choose the range for the variables. This is one of the most important considerations for a number of reasons. First, collecting only a small range of data it might be impossible to distinguish between, say, a linear model, and inverse model, and a power function model. Even a parabolic or hyperbolic function appear linear when a small portion of the curve is considered. Generally speaking, it is best to have as wide a range of data as possible. Determine the extrema, and then consider determining an appropriate number of points between the extrema at which to collect experimental data. Second, data collection is fraught with experimental error, and this error must be minimized to the greatest extent possible. Consider an absolute error (see *Glossary of Terms*) of 1mm. A 1mm absolute error isn't much when one is considering the distance, say, between the floor and the ceiling in a room. The amount of relative error (see *Glossary of Terms*) is small. A 1mm absolute error is huge, for instance, if you are trying the measure the size of a tiny ball bearing. To increase the accuracy of an experiment it is important to reduce the amount of relative error in the data to a minimum.

Step 7. Collect and interpret data. In introductory physics labs this step will generally take the form of graphical analysis (see *Relationships from Graphs*). Using the computer program Graphical Analysis, graph the independent variable against the dependent variable. Perform regression analysis (after linearizing the data if appropriate) and determine the form of the relationship. Pay careful attention to how well your fit matches your theoretical model. If your theory base is correct, then the only reason there should be a difference between the calculated model and the theoretical model is due to random error in the experiment. When you conduct your regression analysis, keep in mind that a 5^{th} order polynomial will accurately fit almost any data. Fortunately, nature is much simpler that that. So, for instance, look for a trigonometric fit rather than a 5^{th} order polynomial fit if you identify an unusual curve. Additionally, be certain that you create a physically reasonable regression model. That is, if the variable on the x-axis is zero and you expect that the variable on the y-axis must similarly be zero, then the regression line must past through the origin (e.g., conduct regression analysis using $y = mx$ instead of $y = mx + b$). Convert the algebraic expression into one that is physically interpreted. Replace y by the variable plotted on the y-axis, and x by the variable plotted on the x-axis. Determine the value and units of the slope and the y-intercept. Give these a physical interpretation if possible.

Step 8. Consider the overall precision of the experiment. If your experimental results in no way match your theoretical model (if used), then there are two sources of possible error. Either your theoretical model is wrong, or you data collection procedures contain a systematic error (see *Glossary of Terms*). You might be asked to use methods dealing with the propagation of error to determine the amount of expected error in a calculated term on the basis of the errors in the experimental terms.