Data Analysis and Generalizations

Exploring Data

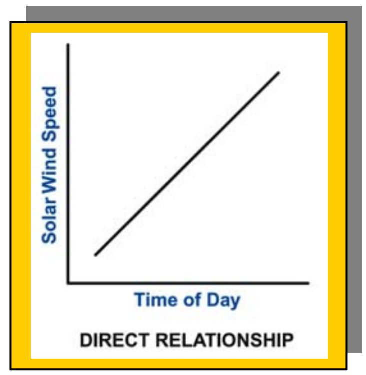
STUDENT TEXT

Scientific Method

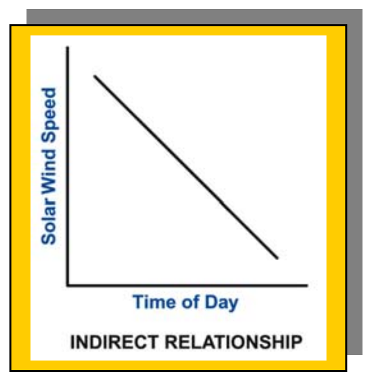
People use the scientific method daily without realizing it. Think about laundry detergent. Why do you buy a particular brand? Your decision is based on your hypothesis. You have a procedure for washing clothes with built-in variables (size of load, temperature of water, amount of detergent, etc.). You run a series of tests (number of loads of wash) until the detergent is gone. By the time you have completed your tests (run out of detergent), you have collected enough evidence to determine the effectiveness of the detergent, and have probably developed a theory.

Observations: A Starting Point

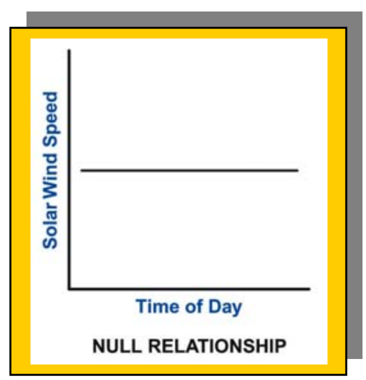
The first step of any scientific investigation is observation of a natural phenomenon or group of phenomena. This is a crucial issue in science. Science begins with observations, whether they be observations of motions of the planets from afar, or observations of the effects of mixing chemical A with chemical B in a laboratory test tube. The observations lead to the hard data of science, which are called facts. A fact is defined by the National Academy of Sciences as "an observation about the character of the natural world that has been repeatedly confirmed." In the activity that you just completed, you were able to make both qualitative (descriptive) and quantitative (numeric) observations of solar wind summary data plots. Observe the speed panel for the solar wind summary data plot for December 19, 2002. A qualitative observation for this panel might be that there was a gradual increase in the speed of the solar wind during this day. So, we can say that as time passed, the solar wind became faster. This is an example of a **Direct Relationship**. On December 20, 2002, the solar wind speed remained about the same. As time passed there was little or no change in the solar wind speed. This is an example of a **Null Relationship**. On December 21, 2002, the solar wind speed gradually decreased. As time passed, the solar wind speed decreased. This is an example of an **Indirect Relationship**. While trends in data can be described as qualitative descriptions, the use of quantitative descriptions is much more precise. Making detailed observations is a good way to learn about the results of your experiment. In this case, you were able to learn about solar wind and the regimes that the Genesis spacecraft experienced over that period of time.



DIRECT RELATIONSHIP



INDIRECT RELATIONSHIP



NULL RELATIONSHIP

You may also wish to **compare** two sets of data to determine whether or not there is **a significant difference** between them. For example, you may have observed that there appear to be changes in speed, temperature, and density when there are changes in solar wind regimes. You have also observed that there are increases and decreases in these characteristics over the period of time that one regime is flowing. These observations may lead you to believe that there are differences in solar wind regime characteristics. You must collect data during a specific period of time and analyze it using the appropriate statistical analysis test before you can determine whether or not there is a **significant** difference, that is, one that has not happened by chance.

Questions and More Questions

Often, good observations lead to questions about the natural world. At the end of the "First Look" activity you listed some questions that you had concerning the solar wind data summary plots. But, not all questions are created equally. First, not all questions may be interesting enough to pursue. Second, you may not have the resources (materials, experience, funding) to perform tests to address these questions. Third, not all questions are stated in ways that are testable. So what makes a good question? The following are points to consider when developing questions:

- A research question should be stated in a way that is testable. For example, questions should not be answered as “yes” or “no.” Rather, begin questions with “To what extent does…” or “How does object A relate to object B?”
- A research question should include a causal relationship and both independent and dependent variables. Variables are defined as anything in the natural world that can exist in varying states or levels that are measurable. The length of daylight on Earth is a good example of a variable. See the Student Text, “Variables and Operational Definitions,” from “*Dynamic Design: Launch and Propulsion*” for further descriptions of variables.
- Do not list all of the variables that you hope to hold constant. While it is important to list these in your research, they do not need to be included in the research question.
- If appropriate, include the area in which you will conduct your research. For example, an unrealistic question might be “How many people in the United States chew brand “X” gum?” It would be nearly impossible to find an accurate answer to this question due to the large scope. A better question might be “How many students at Clear Lake High School chew brand “X” gum?”
- For the purpose of a high school project, research questions should be feasible for student research. For example, it is not reasonable for students to ask a question that might require equipment that is only available in a national research laboratory.

Take a look at the questions that you included in the “First Look” activity. Identify those that might be testable as part of student research for this module. Sometimes, a question may need to be reworded to change it from a poor question to a question that is appropriate for student research.

Hypothesis and Predictions

Formulation of a hypothesis to explain the phenomenon follows the observational, fact-finding phase. The National Academy of Sciences has defined a scientific hypothesis as “a testable statement about the natural world.” Often the statement is of a causal relationship among variables that is testable. While a hypothesis is a limited statement that refers to cause and effect in specific situations, it must offer the opportunity for further testing to be very useful. Students often confuse the terms “hypothesis” and “predictions.” A hypothesis is simply a prediction that includes the tested variables.

Scientists usually write their testable statements in the form of a null hypothesis. For example, if you are going to test whether more girls than boys at Clear Lake High School chew Brand "X" gum, you would write the null hypothesis as: "There is no difference between the number of girls and the number of boys that chew Brand "X" gum at Clear Lake High School." If, after you conduct the study, the data indeed shows that the same numbers of girls as boys chew Brand "X" gum, the data support the null hypothesis. However, if the data show that more girls than boys chew Brand "X" gum, the null hypothesis is not supported. The reason scientists use a null hypothesis is that the null hypothesis provides something that is possible to prove or disprove. In the above example, it is possible to disprove the fact that there is "no difference between the number of girls and the number of boys that chew Brand "X" gum at Clear Lake High School." Disproving that there is a particular difference, such that more girls than boys chew Brand "X" gum, rather than a different difference might not be possible.

A hypothesis can be used to predict the results of observations or the existence of other phenomena. There are two types of predictions that can be made. One is called extrapolation; the other is called interpolation. In the next paragraph, examples of both are provided.

Extrapolated predictions are those that are made outside of the known data points. Trends in the known data can often be used to make accurate extrapolated predictions; however, this is not always the case. Observe the solar wind summary data panel for speed on December 19, 2002, at 12:00:00. Write down the speed at this time. Based on the information presented on this panel for this day, make a prediction for December 20, 2002, at 12:00:00. Write down the speed at this time. Now check the actual information on the panel. How close did you get? This is an example of an extrapolated prediction because it was made outside the known data points.

Interpolated predictions are those that are made between known data points. For an example of an interpolated prediction, recall the actual speed of the solar wind on December 20, 2002, at 12:00:00. Now, look at the solar wind speed on December 22, 2002, at 12:00:00. Write down the speed at this time. Now that you have two data points, predict the solar wind speed on December 21, 2002, at 12:00:00. Now check the actual information on the panel. How close did you get? Was your interpolated prediction more or less accurate than your extrapolated prediction? Typically, interpolated predictions are more accurate. Why do you think this is so?

Once predictions have been made and a hypothesis written, evidence is collected through experimental tests. Several trials should be completed. In the case of the solar wind summary data plots, you will have access to solar wind data like you used in the “First Look” activity for well over 400 days. You will use this vast array of information as you develop an investigation in the next activity.