**Gathering and Analyzing Your Own Data**

**Purpose**

Associations between variables can often be found in data gathered in observational studies. In this project, you will choose variables of interest to you, plan a strategy for gathering appropriate data for those variables, and then examine the data for relationships..

**Directions**

**1.** Choose variables that are of interest to you and for which it is possible for you to gather data by observation or survey methods. Some examples are the following:

* A person's height and how fast he or she runs
* The length of a person's foot and the length of her or his stride
* Grade point average and amount of time watching television
* Attendance in class and period of the day
* Number of hours studying and number of hours working on a part-time job

You or your teacher may want to choose other variables that are more timely for your school or community or more interesting to you.

**2.** Develop a plan for gathering the data. Be sure to make your procedures clear so that your data accurately reflect the variables in which you are interested. Gather a small amount of data from one or two friends or family members. Think about what you know of your test subjects and examine your data to check if the numbers are reasonable. If some of your numbers seem unusual or unexpected, you may need to change your gathering methods. Gather at least 40 pairs of data values.

**3.** Make a table and scatterplot of the data and calculate appropriate summaries.

**4.** Write a report (which is to be a typed paper, PowerPoint or Google Slides, or a poster board – you can also present this report to the class, if you desire) of your study. Include the following, with a justification for each.

* why your variables are of interest
* a description of the steps you took to gather the data and from whom
* the table and scatterplot
* the correlation coefficient (r) and the equation of the regression line (if not appropriate for your data, explain why)
* interpretation of your correlation (strong positive, weak negative, etc.)
* the sum of squared errors
* a discussion of what your results mean about the association of the variables that you studied (including whether strong relationships are cause-and-effect, caused by a lurking variable, or chance)
* the possible effect of influential points on your results (Do you have any? If so, what is that doing for your results? If not, what is that doing for your results?)

**5.** Your presentation and use of the data should reflect the ideas you have learned in this unit. Be sure to make clear in your report what the variables, regression equations, correlations, and other statistics represent.