**Analyzing Existing Data for Association**

**Purpose**

Correlation is everywhere. They are used to help explain how two or more variables are connected; and in turn, this information is used for decision-making in many walks of life. In this project, you will choose an area of interest, research appropriate data, and examine them for possible associations.

**Directions**

**1.** Choose a topic area that is interesting to you and for which data are accessible in references.
Examples are:.

* Your favorite sport, such as softball, baseball, basketball, or tennis
* Popular music, movies, or television shows
* Cars, computers, compact disc players, or other sorts of products

There could be many other topic areas that you might choose, including some that are especially important and timely in your school or community. Your teacher may have some ideas for you, if you have trouble deciding on a topic.

**2.** After you have chosen a topic area, think about variables in your chosen area whose possible association is of interest to you. In the movie industry, what is the relationship between the salary of the star and the profit the movie makes? For cars, how are various ratings related to one another and to the price or sales record of the car? Identify at least three (3) pairs of variables in your topic area whose relationships you will study. (Keep in mind: three variables means three of everything in #4. Four variables means six of everything in #4. Five variables means ten of everything in #4.)

**3.** Find and gather data that you can use in your study. Keep notes describing the sources (books, Internet, people, and so on) of your data. Be sure that you have at least 12pairs of values for each pair of variables you are studying.

**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
* the procedures you used to gather data
* tables and scatterplots of the data for each pair of variables
* the correlation (r) (if not appropriate for your data, explain why)
* interpretation of your correlation (strong positive, weak negative, etc.)
* the equation of a best-fit line (if not appropriate for your data, explain why)
* a discussion of possible outliers (Do you have any? If so, what is that doing for your results? If not, what is that doing for your results?)
* at least one sum of squared errors
* your interpretation of what your results mean in your topic area (including whether strong relationships are cause-and-effect, caused by a lurking variable, or chance)
* a list of references that you used.