This assignment provides you a data file of calibration data for a TCS3200 Color Sensor using an Arduino platform.

The data set consists of RGB color data from Sherwin Williams and a set of readings coming from the sensor under different conditions.

Your goal is to compute regressions for this data using techniques discussed in class.  You should calibrate the individual components separately.  Start with a linear regression and then try to work from there.

In collecting this data, I calibrated it twice.  In the first test, I calibrated it without any modification of the sensor.  These columns are R1, G1, and B1 respectively.  The second test used a black hose washer to provide shielding from external light sources, leaving only the onboard LED available to the readings.  These are R2, G2, and B2 respectively.  You should be comparing R to R1, R to R2, R1 to R2, and the similar comparisons, so this should produce nine models.

The second sheet represents the data collection from a student of mine from a prior summer who worked on spectral interference.  As such, you will find a set of colors connected with the LED (white) on and measures for each individual component with the appropriate light source (red, green, blue).  In those cases, you should try to correlate the readings from actual to the colored light and from LED to the appropriate color.  This should produce six models.

Compare the models using techniques discussed in class.  Which model is better at predicting each of the separate color markers?  Consider residuals and the coefficient of determination in your analysis.

In the second part of your assignment, you also have data about breast cancer cells classified as Benign or Malignant.  Your goal is to determine which measures are the most important indicators for malignancy.  You will need to treat the number of Benign and Malignant as factors.  You should be able to determine if a variable is completely dominant, conditionally dominant, or generally dominant over all others.