Exercise #2 (assigned 2 March, due by 12 March)

Download `exercise2.dat` from the course web page and use it to answer the following questions. Note: For this exercise, I want you to use equations and write things out and solve for things — don't just use canned computer or calculator routines.

(1) The first four columns of `exercise2.dat` contain a two-dimensional data set with errors (that is, $\{x, \sigma_x, y, \sigma_y\}$). Find the best unweighted linear least squares fit to the data. Find the best weighted linear least squares fit to the data.

What are the reduced χ^2 values for these two solutions? What is the covariance of this data set? What is `r` (the correlation coefficient) for this data set? What is the probability that $\{x, y\}$ are correlated?

Briefly discuss your approach and your results.

(2) Now use all six columns of `exercise2.dat` to answer the following questions. You have two data sets with errors here: $\{x, \sigma_x, y, \sigma_y\}$ (columns 1–4) and $\{x, \sigma_x, z, \sigma_z\}$ (columns 1,2 and 5,6). What is the significance of the difference between the means of these two data sets? Use the `t-test` to figure it out. This is a how-many sigma result?

Finally, use the two-sided K-S test to determine of these two datasets are significantly different. If they are not significantly different, what sample size would you need, for the distance (that is, the K-S statistic `D`) that you derive, for the two populations to be significantly different? (You can assume that the two samples always have equal numbers of data points, which need not always be true but here makes the problem a lot easier.) If these two populations are significantly different, what is the smallest sample size that would still find these populations to be different, for the distance (or K-S statistic `D`) that you derive?

Briefly discuss your approach and your results.