

Lecture 2 problems

1. See the following two data files, which are the correlation data of recession velocities and distances from Lemaitre (1927) and Hubble (1929).

Dataset 1: <http://astronomy.swin.edu.au/~cblake/lemaitre.dat>

Dataset 2: <http://astronomy.swin.edu.au/~cblake/hubble.dat>

The object of this exercise is to establish “who discovered the distance-redshift relation”. For each dataset, determine:

- (a) The Pearson product-moment correlation coefficient and its error.
 - (b) The statistical significance of the correlation.
 - (c) Linear least-squares regression lines of the form $V = HD$ and $V = H'D + C$.
 - (d) The errors in H using bootstrap re-sampling.
 - (e) The Spearman rank cross-correlation coefficient and its statistical significance.
2. See the following data file which summarizes two datasets, one with a total of 290 observations, the other with 386 measurements. The former dataset is of flux densities measured at random positions in the sky; the latter is of flux densities at the positions of a specified set of galaxies.

Dataset: <http://astronomy.swin.edu.au/~cblake/twodist.dat>

The object of this exercise is to test whether or not these two samples are statistically consistent.

- (a) Test if the means of the two distributions are consistent.
- (b) Use the Kolmogorov-Smirnov statistic to test if the full distributions are consistent.