ASTR/PHYS 356 - Problem Set 3

Posted: Friday, February 23, 2018 Due: Wednesday, March 7, 2018

Problem 1 – Asteroids (15 pts)
Download the file asteroid_dens.data from the Canvas site
This dataset has 27 rows and the following columns:
Asteroid name
Density in units of g/cm3
Uncertainty of the density (standard deviation)

a) Make a figure where the y-axis shows the name of each asteroid, and the x-axis shows the density. Plot the density of each asteroid in this figure and include the uncertainty in the density as error bars.
c) Plot box and whisker plots for the density measurement and the density error
see: http://matplotlib.org/examples/pylab_examples/boxplot_demo.html
d) What do these numbers and plots indicate?
e) Despite the fact that a sample may appear normal (i.e., Gaussian-like) one can investigate further whether some deviation form normality is present. In this example one might want to do this to verify the possibility that subpopulations of asteroids are present in the sample, due to, for example different structure (solid vs. porous) and composition (rock vs. ice).
One can do this first visually using a normal probability plot (see scipy.stats.probplot). Plot a normal probability plot of the data and describe what you see.
f) One can also use statistical tests to investigate this. Use the following **three** different tests to test the "normality" of the sample:
i- Anderson-Darling
ii- Shapiro-Wilk
iii – Lilliefors test
See Figure 4.7 of "Statistics, Data Mining, and Machine Learning in Astronomy" to see how these tests are used in python.
g) What can you conclude with regards to the existence of subpopulations in this sample?

Problem 2 - Globular Clusters (15 pts)
Use the same data we used in Lab 4 (glob_clus.dat), which can be downloaded from the Canvas site.

This file contains 81 Globular Clusters (GCs) from our Milky Way Galaxy (MW) and 360 GCs from the nearest large spiral galaxy, the Andromeda Galaxy (M31). The first and second columns are GC name and K-band magnitude, respectively. For the MW, *absolute* magnitudes are given. For M 31, *apparent* magnitudes are given. In order to convert the M31 GCs apparent magnitudes to absolute magnitudes (i.e., in order to put them in the same scale as the MW GCs) you need to subtract 24.44 from the apparent magnitude of the M31 GCs.

a) Plot the distributions of both GCs.
b) Estimate the sample mean and the variance for both samples and give their 99% confidence intervals (you can assume that the samples come from a parent Gaussian distribution).
c) Estimate the standard deviation in the mean and the population variance using:
i- the jackknife method
AND
ii- the bootstrap method
(Note: code your own programs that will do these, DO NOT use canned versions of the jackknife and bootstrap methods).
Compare what you get using these resampling methods to their "theoretical" values (assuming the sample comes from a Gaussian parent distribution). Show the histograms of the distributions of mean and variance obtained using these resampling methods and plots of the "theoretical" distributions. See Figure 4.3 of the Astro ML book. **Discuss your findings.**
d) Use the Anderson-Darling and Shapiro-Wilk tests to explore whether the distribution of magnitudes in both the MW and M31 have a Gaussian distribution

Problem 3 – Z statistics (10 pts)
Read about the Z_1 and Z_2 statistics in the Astro ML book (pages 159–160). Note that in these pages they make reference to information and equations shown in Chapter 3 of the book. You should also read that. Explain in your own words what do the Z_1 and Z_2 statistics are good for and why would you want to use them over other tests. Use the Z_1 and Z_2 statistics on the **two** GC samples from Problem 2 and compare with your results in Problem 2-d.
.

Problem 4 – Radio Galaxies (10 pts)
Upload the file "radio_pointings.dat"
This file has two data sets, one with a total of 386 measurements, the other with a total of 290 observations. The former includes measurements of the radio flux density (in mJy) at the position of a specified set of galaxies (detected in the optical); the latter is of flux densities (in mJy) measured at random positions in the sky.
a) plot the data using two histograms using one of the rules of thumb for plotting histograms described in Sec. 4.8.1 of the AstroML book. Indicate which of the rules of thumb you used.
b) Use at least two of the tests we learned in class to test whether there is a statistically significant measurement of radio emission excesses towards the galaxy positions.