AST 475/875 Exercise #3

Due F, September 17th

Masson & Wall (1977; MNRAS 180, 193) examined the luminosity functions of so-called flat and steep radio spectrum quasars to study the cosmological evolution of these objects. The data for their simple source count distributions is available on the course webpage as a text file for both the steep and flat spectrum quasars in their study.

1) Plot the frequency histograms for the steep spectrum and flat spectrum quasars. Comment qualitatively on the relative appearance of these 2 histograms (there really isn't a right or wrong answer for your comments).

2) Plot the cumulative frequency distributions of the two samples, and use a 2-sample Kolmogorov-Smirnov test to address whether the parent distributions of these 2 object classes are the same.

3) Use a Wilcoxon-Mann-Whitney U test to compare the 2 samples and address whether the parent distributions of these 2 object classes are the same. How does this compare with the answer from the K-S test?

Random Notes/Tips:

a) In plotting the histrograms, you have the freedom to choose whatever magnitude bin size you like. Obviously, some judgement/experimentation is needed to choose a bin size which might yield some visually useful information.

b) While you could use the bin values (not the actual data values) to perform the above tests, please use the actual data values. This is a wonderful feature about the K-S and M-W U tests: there is no need to bin your data. Binning loses information: avoid (re)binning, and make use of your actual data whenever you can!!

c) Depending on what software you use for plotting or doing the statistics, you may need to sort your data first into numerical order. Usually this is quite simple to do (even Excel can do it in the blink of an eye).

d) The package SM has, I believe, the ability to construct and graph the cumulative distributions for you with a macro, and to spit out the KS statistic, and assess the probability that the statistic is as large as it is (or larger), but the parent distributions are identical, simply due to chance. I believe SM also will return the wilcoxon-mann-whitney U statistic and assess the probability for you (the assumption is for large N, which is okish here). Ties may not be handled gracefully by SM, however.

e) An annoyance with rank tests is the question of how to handle "ties". I suggest you just parse out ranking for ties as follows. If you had six objects that would take the ranking spots 3,4,5,6,7,8, if their values differed only by teensy weensy amounts, then sum the ranks (33), divide by 6 (to get 5.5), and assign ranking 5.5 to all six objects.