Data Analysis Problems - I

1. Show that the mean of a distribution minimises

$$< (x_i - \hat{x})^2 >_i = \int (x - \hat{x})^2 P(x).dx$$

and that the median of a distribution minimises

$$< |x_i - \hat{x}| >_i = \int |x - \hat{x}| P(x).dx$$

2. In estimating the scatter or spread in a distribution why does the modulus $= < |x - \mu| >$ always lead to lower value than the $rms = \sqrt{< (x - \mu)^2 >}$.

Hint: start from the definition of the two measures and use Schwarz's inequality.

3. For a Gaussian distribution what is the relationship between the modulus, FWHM and MAD measures of the scatter and the more conventional rms (sigma).

4. Two astronomical groups claim independent unbiased estimates of the Hubble constant of 80 ± 15 and 60 ± 10. Combining the two measures, what is the updated best estimate and error. Assume that both error estimates satisfy Gaussian distributions.

Hint: one possibility is to use Bayes' theorem treating one of the measurements as prior knowledge.

If the measurements to be combined were instead 100 ± 15 and 40 ± 10 How would you interpret your result.