STA 302 / 1001 (A. Gibbs)
Sketch of Solutions to Exercises in Chapter 4 of Sheather

1. The data in Table 4.3 are taken from the table at
   `http://www.amstat.org/careers/pdfs/salaryreport_acad2005-6.pdf`
   (the website quoted in the textbook is out of date). You may find it interesting to note that the number of years of experience given in Table 4.3 are estimates from the range of values given in the report.

   SAS code using the counts as weights and the resulting output is given on the web page.

2. Need to find the value for β that minimizes $\sum w_i(y_i - \beta x_i)^2$. Differentiating with respect to β and setting the derivative equal to 0 gives $\hat{\beta} = \frac{\sum w_i x_i y_i}{\sum w_i x_i^2}$.

3. (a) There is a well-known result that, for large n, the variance of the estimate of the median from a sample of size n from a distribution with density function $f(x)$ is $\frac{1}{4n[f(m)]^2}$ where m is median, that is the variance is proportional to 1 over the sample size. Since the data for this question are estimated medians calculated from samples of different sizes, it is appropriate to use weighted least squares with the sample sizes as the weights.

   (b) In the plots of the response versus the predictors, there is a non-linear relationship as well as non-constant variance. From the residual plots, it can be seen that weighted least squares did not solve the problem with non-constant variance, and, of course, there is still a non-linear relationship.

   (c) Try a transformation such as taking the log of the response.