1. Consider a regression model with an intercept, so that the regressors are of the form $\mathbf{x}^T = (1, \mathbf{v}^T)$. Let $\mathbf{C} = \sum (\mathbf{v}_i - \bar{\mathbf{v}}) (\mathbf{v}_i - \bar{\mathbf{v}})^T / (n-1)$ be the sample covariance matrix, and define the Mahalanobis distance by $MD_i^2 = (\mathbf{v}_i - \bar{\mathbf{v}})^T \mathbf{C}^{-1} (\mathbf{v}_i - \bar{\mathbf{v}})$. Show that

$$MD_i^2 = (n-1) \left( h_{ii} - \frac{1}{n} \right),$$

where $\mathbf{H}$ is the hat matrix formed from $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Thus $h_{ii}$ also measures the distance of a regressor from the mean of the data.

2. Let $\theta_F$ be the median of a population with d.f. $F$, and let $\hat{\theta}$ be the sample median. One way to define $\hat{\theta}$ as a functional $h(F_n)$ of the e.d.f. is as $\hat{\theta} = F_n^{-1}(0.5)$. (Look in any mathematical statistics book to see how the inverse of a step function like $F_n$ is defined.) The population version of this is $h(F) = F^{-1}(0.5)$.

(a) Assuming that $F$ has a density $f(x)$ with $f(\theta_F) > 0$, show that the Influence Function of $\hat{\theta}$ is

$$IF(x) = \frac{1}{2f(\theta_F)} \text{sgn}(x - \theta_F).$$

Here $\text{sgn}(\cdot)$ is defined in such a way as to be continuous from the left at 0.

(b) Use (*) to exhibit the asymptotic distribution of the sample median, when sampling from a $\mathcal{N}(\mu, \sigma^2)$ population.

(c) Write $\hat{\theta}$ as an M-estimate, i.e. a solution to $\frac{1}{n} \sum \psi(x_i - \hat{\theta}) = 0$, for a particular choice of $\psi$. Show that the expression for the IF of an M-estimate, derived in class, becomes in this case

$$IF(x) = \frac{\psi(x - \theta)}{E_F[\psi'(X - \theta)]}.$$  

This makes no sense for the $\psi$-function corresponding to the median (why not?). However, first evaluate the denominator for differentiable $\psi$-functions through an integration by parts; and then evaluate it at the $\psi$-function corresponding to the median. Verify that the result agrees with (*).

3. Consider a straight line regression model. Suppose that one of the regressors, say $x_1$, is an outlier which is so far away from the rest of the sample that it is larger than $\bar{x}$, whereas all other $x_i$ are smaller than $\bar{x}$.

(a) Show that the L1 line necessarily passes through $(x_1, y_1)$.

(b) Use (a) to give an alternate proof (i.e. different from the one given in class for monotone M-estimators) that the breakdown point of the estimate is 0.
4. Suppose that $\hat{\theta}$ is an ordinary M-estimate of regression, where scale is known to equal 1 and hence is not estimated. Thus with $r_i = y_i - x'_i\hat{\theta}$ we have

$$\sum_{i=1}^{n} \psi'(r_i) x_i = 0_{p \times 1}.$$  

To judge the influence of the $j^{th}$ data point we might recalculate the estimate after omitting this point from the data, thus obtaining a revised estimate $\hat{\theta}_{(j)}$. This estimate satisfies the equation $F(\hat{\theta}_{(j)}) = 0_{p \times 1}$, where $F(\theta) = \sum_{i \neq j} \psi(y_i - x'_i \theta) x_i$. We can approximate $\hat{\theta}_{(j)}$ by carrying out just one step of the Newton-Raphson method for solving $F(\theta) = 0$, starting with $\hat{\theta}$.

(a) Show that this one-step procedure results in the approximation

$$\hat{\theta}_{(j)} \approx \hat{\theta} - \left[ \sum_{i \neq j} \psi'(r_i) x_i x'_i \right]^{-1} x_j \psi(r_j).$$

(b) Show that the expression above reduces to

$$\hat{\theta}_{(j)} \approx \hat{\theta} - M^{-1} x_j \frac{\psi(r_j)}{1 - \psi'(r_j) x'_j M^{-1} x_j},$$

where $M = \sum_{i=1}^{n} \psi'(r_i) x_i x'_i$.

5. Consider the Water Quality dataset, as described in §5 of SR&C and available on the course website. For each of the methods (i) Least Squares, (ii) Ordinary M-estimation with Huber’s $\psi_{1.5}$, (iii) GM estimation (3 step) with the Huber’s $\psi_{1.5}$ and the ‘w1’ weights, (iv) GM estimation (3 step) with the Huber’s $\psi_{1.5}$ and the ‘w2’ weights, (v) GM estimation (3 step) with the bisquare $\psi_{bi} (; 4.68)$ and the ‘w1’ weights, (vi) MM-estimation, (vii) Least Squares after the removal of the Hackensack River observation:

(a) Fit a regression model relating $N$ to the 3 variables. (Do not use ‘Other’ - fit an intercept instead. Some of the R-functions cannot yet handle no-intercept models.) Present the coefficients in tabular form, so that they can be easily compared. Comment on your findings.

(b) Plot the standardized residuals $r_i/\sigma$ against the fitted values. Comment on your findings.

(c) Carry out individual t-tests for the significance of the three regressors, using the mm fit and appropriate asymptotic approximations.
6. Develop a method of robust non-linear M-estimation. There is no one ‘right’ answer here - I just want you to do something that is sensible, and that agrees with what we have done in class in the linear case. In the model

\[ y_i = f(x_i, \theta) + \varepsilon_i, \]

propose an algorithm for determining an ordinary M-estimate of \( \theta \), i.e. a minimizer of \( \sum \rho \left( \frac{y-f(x_i, \theta)}{\sigma} \right) \) for a suitable function \( \rho \). I suggest using the MAD to re-estimate scale after each of the regression-estimation steps. Show that, if your algorithm converges, then it converges to a solution of the original equations.

7. Suppose that one estimates a straight line for \( x \in [-1, 1] \), obtaining the LS estimate \( \hat{\theta}_0 + \hat{\theta}_1 x \). The design is symmetric, in that \(-x_i\) is a design point whenever \( x_i \) is a design point. Now suppose that the true response is quadratic: \( E[Y|x] = \theta_0 + \theta_1 x + \theta_2 x^2 \).

Define the prediction bias at \( x \) by \( bias(x) = E[\hat{\theta}_0 + \hat{\theta}_1 x] - \{\theta_0 + \theta_1 x + \theta_2 x^2\} \), and the overall bias by \( B = \int_{-1}^{1} bias^2(x) \, dx \). Show that \( bias(x) = \theta_2 (m_2 - x^2) \), where \( m_2 = n^{-1} \sum x_i^2 \), and that

\[ B = 2\theta_2^2 \left\{ \left( m_2 - \frac{1}{3} \right)^2 + \frac{4}{45} \right\}. \]

Then show that the design with equally spaced design points \( x_i = -d + \frac{2(i-1)}{n-1} \), where \( d = \sqrt{\frac{n-1}{n+1}} \), is a bias-minimizing design.