We consider the generalization of multiple linear regression involving \( l \) predictor variables \( x_1, x_2, \ldots, x_l \) and \( k \) (unknown) parameters \( \beta_1, \beta_2, \ldots, \beta_k \) (that, as convenient, we will assemble into vectors \( x \) and \( \beta \) respectively). We assume that there is some known function \( f(x; \beta) \) that provides the mean value of an observable variable, \( y \), in terms of these. Then, as in MLR, we assume that for independent mean 0 normal variables \( \varepsilon_i \), for \( i = 1, 2, \ldots, n \)

\[
y_i = f(x_{i1}, x_{i2}, \ldots, x_{il}; \beta_1, \beta_2, \ldots, \beta_k) + \varepsilon_i
\]

(*)

Notice that in model (*), exactly as in MLR, there is an assumption that for inputs \( x_1, x_2, \ldots, x_l \) the distribution of \( y \) around the mean \( \mu_{y|x_1, x_2, \ldots, x_l} = f(x_1, x_2, \ldots, x_l; \beta_1, \beta_2, \ldots, \beta_k) \) is normal with a standard deviation \( \sigma \) that doesn't depend upon the inputs. The innovation here as compared to MLR is simply the possibility that \( f \) doesn't have the MLR form \( \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \).

Particularly in the physical sciences and engineering theories e.g. involving differential equations and well-established empirical relationships provide other favorite functional forms. Some of these don't even have explicit forms and are simply defined as "the solution to a set of differential equations." This handout is about how they can to some extent be handled statistically in a way parallel to the handling of MLR.

First, there is the question of how to process \( n \) data vectors into estimates of the parameters \( \sigma \) and \( \beta \). Just as in MLR, one can use least squares, i.e. minimize

\[
Q(\beta) = \sum_{i=1}^{n} (y_i - f(x_{i1}, x_{i2}, \ldots, x_{il}; \beta))^2
\]

Conceptually, this exactly as in MLR (except for the fact that the function \( Q(\beta) \) is not a quadratic function of \( \beta \)). Operationally (because \( Q(\beta) \) is not so simple) one must employ iterative algorithms to search for an optimizer of \( Q(\beta) \). In \( \mathbb{R} \) one must use the \texttt{nls()} routine instead of the \texttt{lm()} routine. Further, because statistical theory for the general model (*) is not as clean as for the special case of MLR, only approximate methods of inference can be identified and they are impossible to describe completely at this level of background. What will be done in the balance of this handout is to try to give some understandable description of and motivation for what is possible and is implemented in good statistical packages.

So, without getting into the numerical details of exactly what algorithms are used to locate it, suppose that \( \mathbf{b} \) is an optimizer of \( Q(\beta) \), that is
\[ Q(b) = \min_{\beta} Q(\beta) \]

(\(b\) minimizes an "error sum of squares") and is an "ordinary least squares estimate" of \(\beta\). Of course, entries of \(b\) serve as estimates of the entries of \(\beta\).

The (minimum) sum of squares corresponding to \(b\)

\[ SSE = Q(b) = \sum_{i=1}^{n} \left( y_i - f(x_{i1}, x_{i2}, \ldots, x_{ik}; b) \right)^2 \]

can (as in MLR) form a basis for estimating \(\sigma\). In particular a simple estimate of \(\sigma^2\) is

\[ \hat{\sigma}^2 = \frac{SSE}{n-k} \]

Simply carrying over ideas from MLR, a very approximate set of confidence limits for \(\sigma\) are

\[ \hat{\sigma} \sqrt{\frac{n-k}{\chi^2_U}} \quad \text{and} \quad \hat{\sigma} \sqrt{\frac{n-k}{\chi^2_L}} \]

(for \(\chi^2\) values based on \(n-k\) degrees of freedom). More subtle methods than this are available and are sometimes available in non-linear least squares software.

Confidence regions for locating the whole parameter vector \(\beta\) are sometimes of interest. Carrying over an idea from MLR (that wasn't discussed in the context of MLR does work there exactly) one can use as a confidence region a set of parameters \(\beta\) for which \(Q(\beta)\) is not much larger than the minimum value, \(Q(b)\). In particular, the set of parameters for which

\[ Q(\beta) \leq Q(b) \left( 1 + \frac{k}{n-k} F_{U} \right) \]

serves as an approximate confidence region (with confidence level the probability that the F distribution with degrees of freedom \(k\) and \(n-k\) places below the value \(F_{U}\) in the display above).

Standard errors for the \(b_j\) can be obtained in much the same way as they are in logistic regression, based on the curvature of an appropriate log-likelihood function (involving the inversion of a Hessian, etc.). These values \(SE_{b_j}\) are printed out on most non-linear regression outputs and approximate confidence limits for \(\beta_j\) are
\[ b_j \pm t \cdot SE_{b_j} \]

t an upper percentage point for the \( t_{n-k} \) distribution.

More reliable approximate confidence limits for individual coefficients \( \beta_j \) can be made via a method related to the method for making joint confidence regions for \( \beta \) discussed above. That is, if one identifies the set of parameters for which

\[ Q(\beta) \leq Q(\hat{\beta}) \left( 1 + \frac{1}{n-k} F_U \right) \]

(for \( F_U \) an upper percentage point of the F distribution with degrees of freedom 1 and \( n-k \)) and finds the smallest "box" containing it, say the set of \( \beta \) with

\[ L_j \leq \beta_j \leq U_j \quad \text{for } j = 1, 2, \ldots, k, \]

then the intervals \( (L_j, U_j) \) function as individual confidence limits. (For example, the upper F 5\% point gives individual 95\% intervals.) This method is sometimes implemented in non-linear regression programs as an alternative to the t intervals.

It is also possible to make confidence limits for the mean response \( f(x; \beta) \), but no version of this seems to be presently implemented in \texttt{nls()} . In particular, no standard errors for fits (the \( SE_{\hat{y}} \)) seem to be implemented at the moment. If they were, then approximate prediction limits for a next \( y \) at a particular set of conditions would be

\[ f(x; \hat{\beta}) \pm t \sqrt{\sigma^2 + (SE_{\hat{y}})^2} \]