This exercise sheet is centered around a data set from an investigation concerning an insulin-like growth factor (IGF-1) in a group of 1339 healthy humans, primary school children. The variables are age, the age of the person in years, \texttt{menarche}, indicating whether a girl has had her first period (1: no, 2: yes), \texttt{sex} (1: boy, 2: girl), \texttt{growth}, insulin-like growth factor in \textit{$\mu$g}/l, \texttt{stage}, the tanner stage of puberty classified from 1 to 5, and \texttt{testicular}, the testicular volume in \textit{ml}.

(a) Import the data set and, to get a first impression of the data, perform a plot of the insulin-like growth factor as a function of age (you might use \texttt{read.table}).

(b) For persons of more than 25 years one might expect the following behavior, \texttt{growth} $\sim \sqrt{\texttt{age}}$, i.e. \texttt{growth} = $a\sqrt{\texttt{age}} + b$, for \texttt{age} $\geq 25$. Perform the corresponding regression analysis and comment on your result (you might use \texttt{subset} in \texttt{lm}).

(c) Now we look at the subset of the data for persons of age less than 25 years (use \texttt{subset}). For this subset analyze the linear dependence of \texttt{growth} on \texttt{age} with interactions with the factors \texttt{sex} and \texttt{stage} by plotting the $2 \times 5$ regression lines. Also try transformations of the dependent variable \texttt{age}, i.e. log and square root. Use the regression diagnostics to decide which fits the data best.

(d) Perform a logistic regression, i.e. a generalized linear model (\texttt{glm}) with link function logit $p = \log[p/(1-p)]$, of \texttt{menarche} with dependent variable \texttt{age}. Plot the fitted probability of menarche having occurred and determine the median menarcheal age.