1. Condsider the linear model $\mathbf{Y} \sim \mathsf{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n)$ with $\boldsymbol{\mu} = \mathbf{1}\beta_0 + \mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}$ a full rank matrix with rank $p$

   (a) Show that the projection, $\mathsf{P}$, on the column space spanned by $\mathbf{1}$ and $\mathbf{X}$ may be written as $\mathsf{P} = \mathsf{P}_{\mathbf{1}} + \mathsf{P}_{\mathbf{X}-\mathbf{1}\bar{\mathbf{x}}}$. Show that diagonal elements are

   $$h_{ii} = \frac{1}{n} + (\mathbf{x}_i - \bar{\mathbf{x}})^T \left( (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}})^T (\mathbf{X} - \mathbf{1}\bar{\mathbf{x}}) \right)^{-1} (\mathbf{x}_i - \bar{\mathbf{x}})$$

   (Recall all vectors are column vectors). The $h_{ii}$ are known as the leverage values.

   (b) Find the sampling distribution of $\hat{\mu}_i$ (the mean of $Y_i$ at $\mathbf{x}_i^T$ as a function of $h_{ii}$ and provide an expression for a 95% confidence interval. For what values of $\mathbf{x}$ will the interval be the narrowest? Explain.

   (c) Given $\sigma^2$, find the distribution of $\mathbf{e}_i$ as a function of $h_{ii}$. Explain (rigorously) why $\mathbf{e}_i$ unconditional on $\sigma^2$ does not have a student $t$ distribution with $n - p - 1$ degrees of freedom.

   (a) Refer to the Prostate data from `library(lasso2); data(Prostate)`

   (b) Fit a linear model using `lcavol` (log cancer volume) as the response and include all covariates. Construct 95% confidence intervals for each coefficient and provide a meaningful interpretations for changes in the cancer volume ( not log cancer volume) include any units etc in your interpretation. See Wakefield page 1.3.1 for details on variables. Note "a 1 unit" change may or may not be meaningful for interpretation.

   (c) Fit the regression model with response `lcavol`, and variables `svi` and `lpsa` as predictors. Plot the cancer volume versus PSA on the log scale. Add the fitted regression function for svi = 1 and svi = 0, with lines representing the (pointwise) 95% confidence intervals for each.