1. Add 95% prediction intervals to your plot from HW3 for the Prostate data using a different linetype and color. Explain why the prediction intervals are wider than the confidence intervals for $\hat{\boldsymbol{\mu}}$. (See the function `predict()` in R. Please label all axes with units and informative names, add a legend to explain the multiple lines, and a caption).

2. Consider the linear model 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. For a new observation Y_* at $\mathbf{x}_*$ with $Y_* = \mathbf{x}_*^T \boldsymbol{\beta} + \epsilon_*$ and ϵ_* independent of $\boldsymbol{\epsilon}$, consider the predicted residual $Y_* - \mathbf{x}_*^T \hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ is the MLE using data $\mathbf{Y}$.

   (a) Find the distribution of the predicted residual $Y_* - \mathbf{x}_*^T \hat{\boldsymbol{\beta}}$ given $\boldsymbol{\beta}$ and σ^2.

   (b) Show that the standardized predicted residual (center so that the mean is 1 and and scale (sd) is 1 with σ^2 replaced by the usual unbiased estimate $\hat{\sigma}^2 = \mathbf{Y}^T(\mathbf{I} - \mathsf{P}_{\mathbf{X}})\mathbf{Y}/(n - p - 1)$ has a student t distribution. What are the degrees of freedom?

3. Consider the linear model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ with $\mathsf{E}[\boldsymbol{\epsilon}] = \mathbf{0}_n$ and $\mathsf{Cov}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$ and with $\mathbf{X}$ of full column rank $(p + 1)$.

   (a) Consider estimation of $\boldsymbol{\beta}$ using quadratic loss $(\boldsymbol{\beta} - \mathbf{a})^T(\boldsymbol{\beta} - \mathbf{a})$ for some estimator $\mathbf{a}$. Find the expected quadratic loss if we use the MLE $\hat{\boldsymbol{\beta}}$ for $\mathbf{a}$. Simplify the expression as a function of the eigenvalues of $\mathbf{X}^T\mathbf{X}$. What happens as the smallest eigenvalue goes to 0?

   (b) Consider estimation $\boldsymbol{\mu}$'s at the observed data points $\mathbf{X}$. Find the expected quadratic loss $\mathsf{E}[(\boldsymbol{\mu} - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\boldsymbol{\mu} - \mathbf{X}\hat{\boldsymbol{\beta}})]$. What happens as the smallest eigen value of $\mathbf{X}^T\mathbf{X}$ goes to 0?

   (c) Consider predicting $\mathbf{Y}_*$'s at the observed data points $\mathbf{X}$ where $\mathbf{Y}_*$ is independent of $\mathbf{Y}$. Find the expected quadratic loss $\mathsf{E}[(\mathbf{Y}_* - \mathbf{X}\hat{\boldsymbol{\beta}})^T(\mathbf{Y}_* - \mathbf{X}\hat{\boldsymbol{\beta}})]$. What happens as the smallest eigen value of $\mathbf{X}^T\mathbf{X}$ goes to 0?

   (d) Consider predicting $\mathbf{Y}_*$'s at new points $\mathbf{X}_*$ with $\mathsf{E}[\mathbf{X}_*^T\mathbf{X}_*] = \mathbf{I}_p$. Find the expected quadratic loss $\mathsf{E}[(\mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}})^T(\mathbf{Y}_* - \mathbf{X}_*\hat{\boldsymbol{\beta}})]$. What happens as the smallest eigen value of $\mathbf{X}^T\mathbf{X}$ goes to 0? (If $E[\mathbf{X}_*^T\mathbf{X}_*] = \boldsymbol{\Sigma} > 0$ does that change the result)

   (e) Comment on the difference in estimation, prediction at observed data and prediction at new data as $\mathbf{X}$ becomes non-full rank. Which is the most stable? Which is the least?