1. Faraway (1st edition) problem 5.4

2. Faraway (1st edition) problem 5.5

3. Using the same covariates and response as in Faraway 5.4, use bootstrapping to check whether the usual estimate of the standard error for $\hat{\beta}_{\text{GNP.deflator}}$ seems to estimate the variability appropriately. (Bootstrap the sample, not the residual.) Discuss what you see and any possible explanations.

4. In this problem we'll prove that, no matter the correlation structure of the covariates, variance of estimating a mean at a new $x \in \mathbb{R}^p$ can only increase when you add an additional covariate. You are welcome to collaborate in pairs or groups of three on this problem; if you choose to work in a group, please list your collaborators in your handed in HW.

   We will consider two models: with and without X_j. Let $X \in \mathbb{R}^{n\times p}$ be the full matrix of covariates and X_{-j} be the same matrix with the X_j column removed. We will assume that the normal linear model holds in both cases, i.e. the true model for the response is

$$Y_i = \beta_1 X_{i1} + \cdots + \beta_p X_{ip} + N(0, \sigma^2)$$

   and we have $\beta_j = 0$ so that this model is true even with X_j removed. We'll write $\hat{\beta}$ for the fitted coefficients using all the covariates, and $\hat{\beta}_{-j}$ for the model using the $p-1$ covariates when X_j is removed. Note that $\hat{\beta}_{-j}$ is not the same as removing the entry j from the vector $\hat{\beta}$—the values may have changed entirely.

   If we predict the mean response at a new $x \in \mathbb{R}^p$, we would predict

$$\hat{y} = x_1\hat{\beta}_1 + \cdots + x_p\hat{\beta}_p = x^\top\hat{\beta}.$$

   For the reduced model, we would predict

$$\hat{y}_{-j} = x_1(\hat{\beta}_{-j})_1 + \cdots + x_{j-1}(\hat{\beta}_{-j})_{j-1} + x_{j+1}(\hat{\beta}_{-j})_{j+1} + \cdots + x_p(\hat{\beta}_{-j})_p = x_{-j}^\top\hat{\beta}_{-j}$$

   X_j 0

   where x_{-j} is the vector x with entry j removed.

   We will use a linear algebra result:

   Lemma: For any positive definite matrix $\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$, it holds that

$$\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}^{-1} \succeq \begin{pmatrix} A^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

   Here $M \succeq N$ is the positive semidefinite ordering, defined on matrices M, N which are themselves positive semidefinite, with $M \succeq N$ equivalent to $M - N \succeq 0$, i.e. $M - N$ is positive semidefinite.

   You can use the fact that, for positive semidefinite and invertible M, N, it holds that $M \succeq N$ if and only if $M^{-1} \preceq N^{-1}$.

   (a) Write down the variance of $\hat{y}$ and of $\hat{y}_{-j}$, using matrix notation such as $X^\top X$ for short and clean answers.

   (b) Consider predicting the mean response value at a new $x \in \mathbb{R}^p$. Assuming the lemma is true, prove that $\text{Var}(\hat{y}) \geq \text{Var}(\hat{y}_{-j})$.

   (c) Prove that if X_j is orthogonal to X_k for every $k \neq j$, and $x_j = 0$, then the variances are in fact equal.

   (d) Now we'll prove the lemma. One simple way to do this is with a limiting argument—we'll prove that

$$\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}^{-1} \succeq \begin{pmatrix} \frac{1}{1+\epsilon}A^{-1} & 0 \\ 0 & 0 \end{pmatrix}. \qquad (*)$$

   If this is true for any $\epsilon > 0$ then taking a limit, the lemma will be true.

   We'll break the proof into steps:

i. Prove that if $\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$ is positive semidefinite, then so is $\begin{pmatrix} \epsilon A & -B \\ -B^\top & \epsilon^{-1}C \end{pmatrix}$.

ii. Using the previous step, prove that $\begin{pmatrix} \epsilon A & -B \\ -B^\top & c\mathbf{I} - C \end{pmatrix} \succeq 0$ for a sufficiently large constant c (you should specify c in terms of the other quantities in the problem).

iii. Using the previous step, prove that this implies $\begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}^{-1} \succeq \begin{pmatrix} \frac{1}{1+\epsilon}A^{-1} & 0 \\ 0 & c^{-1}\mathbf{I} \end{pmatrix}$ and that this implies the equation marked with a (*) above.