1. In `gala` dataset, fit a multiple linear regression model with Species as response variable, and all others as predictors.

   (a) Use `lm()` command and show me the R output.

   (b) Write out the formula to estimate $\widehat{\text{Species}}$.

   (c) Estimate the value of Species if Endemics is 18, Area is 2.6, Elevation is 190, Nearest is 3.05, Scruz is 81, and Adjacent is 50.

   (d) In R, use `solve()`, `t()`, $\% * \%$, `cbind()`, `rep()` to find $\hat{\beta}$ directly **WITHOUT** using `lm()`. (*Hint: Check slides page 18.*)

2. In `gala` dataset, fit a multiple linear regression model with Species as response variable, and all others as predictors.

   (a) Find the value of $\hat{\sigma}$ in `summary(fit)` directly, and calculate $\hat{\sigma}^2$.

   (b) Use the method in slides page 22 to calculate $\hat{\sigma}^2$, do you get the same value in (a)?

   (c) Use the formula

   $$\hat{\sigma}^2 = \frac{\mathbf{Y}^T(\mathbf{I} - \mathbf{H})\mathbf{Y}}{n - p - 1}$$

   and matrix/vector calculation method in R to calculate $\hat{\sigma}^2$. Note that

   $$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$$

   and the R command to create a 30×30 identity matrix $\mathbf{I}$ is `I <- diag(rep(1,30))`.

   (d) Calculate the 95% confidence interval for **Endemics** and **Adjacent**, and make solid interpretation.

   (e) Read the p-value of **Endemics** and **Adjacent**, make your decision and write interpretation. Are the testing results similar to the one in (d)?

3. In `gala` dataset:

   (a) Use Backward Elimiation to select the best subset of the predictors. Interpret propoerly for the estimated β's in your final model.

   (b) Use Forward Selection to select the best subset of the predictors.