STAT 636, Fall 2017 - Assignment 5
UNGRADED - Practice for Exam II

Consider the `happiness` data. Let the response Y be `Distopia`, and consider the numeric variables `Economy`, `Family`, `Health`, `Freedom`, `Trust`, and `Generosity` as the predictor variables; call these $x_1, x_2, \ldots, x_6$, respectively. For the purpose of predicting the value of Y, we will use linear regression models of the form

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_6 x_{6i} + \epsilon_i$$

where the ϵ_i are IID $N(0, \sigma^2)$, $i = 1, 2, \ldots, n$.

1. Using a usual least squares linear regression model (the `lm` function in `R`):
   (a) Use leave-one-out cross-validation to estimate the MSE of your model.
   (b) Use bootstrap to estimate the standard deviation of your MSE estimate.
2. Using regularized (lasso-based) linear regression (the `glmnet` and `cv.glmnet` functions from `glmnet` package in `R`, with `family='gaussian'`):
   (a) Based on cross-validation, using the `cv.glmnet` function, what is the optimal value of the tuning parameter `lambda`?
   (b) Use leave-one-out cross-validation (code it up yourself) and `glmnet` to estimate the MSE of the lasso model using the optimal tuning parameter.
   (c) Use bootstrap (again, code yourself) to estimate the standard deviation of your MSE estimate.
   (d) Compare the estimated β coefficients from the lasso model to the least-squares model, and confirm that the lasso model's coefficient estimates have been "shrunken" toward 0.