STAT342 Homework 1

Due: Friday January 22, 4:00 pm in the STAT 342 folder in the homework cart near the door of Statistics 102.

Note: Boldface implies vector.

There is a file `cholDat.RData` posted on Canvas which contains data on the level of cholesterol (in mg per liter) found in the bloodstream of 10 Olympic athletes. The average daily intake of saturated fat (in mg) is also given. Treat cholesterol as the response variable (y) and saturated fat as the covariate (x).

1. Using the `lm` tool in R, fit a simple linear regression model $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$; $E[\epsilon_i] = 0$ and $Var[\epsilon_i] = \sigma^2$ to this data. Report *and interpret* your estimates $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}$. Report units on the parameter estimates where it makes sense to do so.

2. Is the relationship between saturated fat and cholesterol significant? Explain.

3. Recall that the simple linear model can be written in matrix form: $\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Write down what the design matrix X would be for this data set. (You don't have to write down all the entries, but give enough information so that you clearly convey that you know what the design matrix should be.)

4. Find the estimates $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\sigma}$ by using `R` to solve the matrix form of the problem.

5. Is this simple linear model adequate for this data? If your answer is "yes", then defend this model. If your answer is "no" then provide support for your conclusion. Recall the diagnostics from linear regression, as these will likely help you to make your argument. If your answer is "no", suggest an alternative model.

6. Like nearly all statistical analyses, this data represents a sample from some population. Given what you know about this sample, what population would it make sense make inference on? That is, for what population do you think you could apply the model you fit?