PROBLEM 1

Researchers are interested in providing doctors an easy way to predict a patients body fat percentage from physical measurements. For 252 men, they accurately measure their percentage of body fat with an underwater weighing technique and additionally collect their weight (in pounds) and 10 body circumference measurements, including abdominal circumference (in cm).

One model the researchers consider is:

$$fat_i = \beta_0 + \beta_1(weight_i - 178.9) + \beta_2(abdom_i - 92.5) + \epsilon_i \quad i = 1, \ldots, 252$$

where,

- *fat* is the patient's fat percentage,
- *weight* is the patient's weight in pounds, and
- *abdom* is the patient's abdominal circumference in cm.

178.9lbs and 92.5cm are the average values of weight and abdominal circumference in the participants studied.

Some details of the least squares model fit are given below:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} 19 \\ -0.14 \\ 0.92 \end{pmatrix}, \qquad \hat{\sigma}^2(X^TX)^{-1} = \begin{pmatrix} 0.07 & 0 & 0 \\ 0 & 0.0004 & -0.0009 \\ 0 & -0.0009 & 0.003 \end{pmatrix}$$

A. [2 points] What assumptions are needed to derive the form of least squares estimate of $\beta = (\beta_0, \beta_1, \beta_2)$?

B. [6 points] Construct a 95% confidence interval for β_1 and state the assumptions necessary for the interval to be exact.

C. [4 points] Interpret your interval in part (B) in the context of the problem.

D. [2 points] What is the estimated mean fat percentage for a man with a weight of 177.9lbs and an abdominal circumference of 93.5cm?

E. [2 points] How would you construct a confidence interval for the estimate in part (D)? (Be specific about what you would calculate, but do not do the calculation)

F. [2 points] When would it be **more appropriate** to report a prediction interval for the estimate in part (D)?

G. [2 points] The researchers also consider including all 10 body circumference measurements in their model. Give a reason why it may be **useful** to perform variable selection on the 10 body circumference measurements.