Dear all, this homework is due before class on Tuesday, February 27, 2024. Please submit it to SBU Brightspace. In this homework, we shall work on linear regression analysis using the famous Galton data on the heights or parents and their children. This is the original study where the term "regression" comes from. You can read the original paper from Galton here: http://galton.org/essays/1880-1889/galton-1886-jaigi-regression-stature.pdf

1. **Please establish a new variable 'midparent' that is the average height of the mother and father pair. Please answer the following questions using the least squares method and hand calculation.**
    a. **Please establish a simple linear regression model to predict daughter's height using the midparent height;**
    b. **Please establish a simple linear regression model to predict son's height using the midparent height.**

2. **Please establish one general linear model (namely, multiple regression with at least one categorical predictor, here being gender of the child), to predict the child's height, one model for both the daughters and the sons, using the following predictors: mother's height, father's height, child's gender. You can use either hand calculation or R pogram, or Python program – please attach your code when submitting your homework solution if you use R or Python programming.**

3. **Which method (the simple linear regression method in Question 1a, or the multiple regression method in Question 2) is better in predicting daughter's height? Why? (Hint: compare the sum of squared error [SSE] or the mean squared error [MSE], for example.)**

4. **Which method (the simple linear regression method in Question 1b, or the multiple regression method in Question 2) is better in predicting son's height? Why? (Hint: compare the sum of squared error [SSE] or the mean squared error [MSE], for example.)**