CS 301
Module 4 Homework

1. A lab would like to know if there is a significant relationship between the amount of nitrogen (in lbs) applied to fields and the wheat yield (in bushels). The data are shown below. Answer the questions below using calculations on an electronic spreadsheet. Do not use any python regression algorithms. Show your work. **(30 points)**

Nitrogen (x)	0	0	20	20	40	40	60	60	80	80	100
Yield (y)	41	54	127	149	348	299	419	483	481	619	549

   a. **Find the best fitted line of the data given above**
   b. **Find the R-squared value**
   c. **Find the F value of the best fitted line**

$$F = \frac{\left(SS_{mean} - SS_{fit}\right) / (p-1)}{SS_{fit} / (n-p)}$$

   d. **Predict the expected yield if 50 pounds of nitrogen are applied.**

2. Assume we gathered a random sample of the following dataset, where the independent variable (x) represents the score (out of 40) a student achieved on a standard math exam, and the dependent variable represents the score (out of 40) of the student on a Data Science exam. Is there a correlation between the two variables, and if so, how strong is this correlation? Your calculations should be done using an electronic spreadsheet. Show your work. **(15 points)**

Math	20	23	8	29	14	11	11	20	17	17
DS	30	35	21	33	33	26	22	31	33	36

3. A population has a mean of 40 and a standard deviation of 15. A random sample of size n = 100 is selected. What is the probability that the sample mean is between 38 and 44? Hint: need to use the concept of the normal distribution and z-score. Python (and appropriate libraries can be used to answer this question. Show your work. **(15 points)**