

Assignment 1: Linear Regression

Due date: September 22, 11:59pm

Attention: Please prepare two files for each homework assignment: a .pdf file for your answers including relevant figures, and a .R file for your relevant R scripts. File names should be `Last_First_hw.pdf` and `Last_First_hw.R`, e.g., `Leonard_Nemoy_1.pdf` and `Leonard_Nemoy_1.R`. Your submissions must be based on your own original work. Late submissions will not be accepted.

A farmer is trying to optimize the crop yield by varying the amount of water, fertilizer and herbicide used. The file `Crops.csv` contains data on various experiments the farmer conducted, documenting the amount of each input used and the resulting crop yield. Crop yields are also affected by other attributes that cause the yield to randomly vary.

1. Load the data from `Crops.csv` into R. You can use `setwd()` to set the current working directory. Print a summary of the variables.
2. Regress the yield on the amount of water used. Explain and interpret the results.
3. Regress the yield on the amount of fertilizer used. Explain and interpret the results.
4. Regress the yield on the amount of herbicide used. Explain and interpret the results.
5. Regress the yield on all the variables. Explain and interpret the results.
6. The farmer suspects that high levels of fertilizer may not be effective. To check this conjecture, plot the yield against the amount of fertilizer used. Explain why the plot is consistent with the regression results.
7. Based on the plot, create an indicator `appropriateFertilizer` whose value is 1 when the amount of fertilizer is appropriate, and 0 when the amount of fertilizer is too high or too low. Regress Yield on the indicator you created and interpret the results.
8. The farmer suggests that an appropriate amount of fertilizer should raise the effectiveness of watering the crops. Run a regression with an interaction between water and `appropriateFertilizer` to check this. Interpret the results.
9. Select a collection of variables and interaction terms to use as predictors for the yield. Run the regression, and interpret the results. Explain why you chose this regression model.

10. For this model, what is a 99% confidence interval for the regression coefficients. Interpret the results.
11. For this model, what is a 90% prediction for the yeild of a single sample that gets water=30,fertilizer =5 and herbicide =5? Interpret the results.