GEB 6895: Business Intelligence
Department of Economics
College of Business Administration
University of Central Florida
Fall 2019

Assignment 4

Due Tuesday, October 16, 2019 at 11:59 PM
in *your* private mirror of the GEB6895F19 GitHub repo.

Instructions:

Complete this assignment within the space on your private mirror of the GEB6895F19 GitHub repo in the folder `assignment_04`. Create a folder called `my_answers` that will contain all of your work for this assignment. Within this folder, code your solutions in `.R` with the filename as specified. When you are finished, use `git` to `add`, `commit` and `push` your code to your private mirror of the GEB6895F19 repo. You are free to discuss your approach to each question with your classmates but you must `git push` in your own work.

Question 1:

In this exercise, you will produce a script that calculates the OLS estimator for a linear regression model, using a number of numerical methods. Use the script `Calculating_beta_hat.R` and save it in your folder called `my_answers` in the folder `assignment_04`.

The script provides estimates of the model $y = \beta_0 + \beta_1 x$, where y is aggregate income and x is the percentage of the labor force employed in agriculture.

a) Obtain the value of the slope coefficient on the variable `agg_pct`, available from the object from the function `coef(lm_model)` and store it as `beta_1_hat_lm`.

b) Calculate the slope coefficient using the OLS estimator

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

where $\bar{x}$ is the mean of x and $\bar{y}$ is the mean of y. Store this value as `beta_1_hat_calc`.

c) Obtain another estimate of the OLS estimator by solving a system of normal equations

$$X^T y = X^T X \hat{\beta},$$

which is, in matrix form,

$$\begin{bmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} X_i y_i \end{bmatrix} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

i) Create the two matrices that represent the system of normal equations.

ii) Use the `solve` function to solve the system of normal equations and obtain another estimate for $\hat{\beta}_1$. Store this value as `beta_1_hat_norm`.

d) Estimate $\hat{\beta}_1$ by minimizing the sum of squared residuals, defined as

$$SSR(\beta) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x)^2$$

i) Create a function `ssr(beta, y, x)` that returns the value of the expression $SSR(\beta)$. You can test your function by reproducing the value of `Residual standard error` with the value $\sqrt{SSR(\beta)/8}$.

ii) Plot the SSR function on a graph by drawing a few lines for fixed values of β_0 with β_1 varying across the horizontal axis. Make sure one of the lines corresponds to the estimated intercept coefficient $\hat{\beta}_0$.

iii) Use the `optimize` function to minimize $SSR(\beta)$ and obtain another estimate for $\hat{\beta}_1$. Store this value as `beta_1_hat_opt`.