QMB 6316: R for Business Analytics
Department of Economics
College of Business
University of Central Florida
Fall 2024

Assignment 2

Due Sunday, November 17, 2024 at 11:59 PM
in digital form on Webcourses.

Question 1:

This activity follows the in-class RStudio exercise using the script `OLS_sim.R`, which we will extend with the script `OLS_omit_A2.R` in this exercise. The objective is to show that you can successfully run an R program in RStudio, obtain statistical results, modify that program, and obtain revised results. This exercise simulates the situation in which you are responsible for drawing conclusions from a script and dataset provided by someone else.

In this exercise, we will generate simulated data for the prices of cars, which depend on mileage, whether the car has been in an accident, and whether the car has sustained major structural damage, which can happen only as a result of an accident. Specifically, the regression model is

$$CAR\_PRICE_i = \beta_0 + \beta_1 \times MILEAGE_i + \beta_2 \times ACCIDENT_i + \beta_3 \times DAMAGE_i + \epsilon_i \qquad (1)$$

where:

CAR_PRICE_i	=	the value of car i
$MILEAGE_i$	=	the mileage of car i
$ACCIDENT_i$	=	whether or car i has been involved in an accident (i.e., $ACCIDENT_i = 1$ if car i has been in an accident, zero otherwise)
$DAMAGE_i$	=	whether or not car i car has sustained major structural damage (1 or 0)

Run the entire script and compare the output from `summary(lm_full_model)`, which includes all variables, with that from `summary(lm_no_damage)`, which omits the damage indicator. If there are no cars with damage in your simulation, run the script again to take another draw.

a) Copy and paste the regression model estimates after the commands `summary(lm_full_model)` and `summary(lm_no_damage)`.

b) Compare the estimated coefficient for $ACCIDENT_i$ with and without the damage variable. How does this relate to the coefficient for $DAMAGE_i$?

c) Compare the values of R^2 (labeled `Multiple R-squared`) and $\bar{R}^2$ (labeled `Adjusted R-squared`) for the two models. Which model do you recommend (pretending that you don't know the true model)?

Now consider the situation in which car drivers have perfect insurance against the damages caused in an accident. That is, any damages from accidents are fully repaired with no decrease in the value of cars. To implement this in the script `OLS_omit_A2.R`, change line 86

from `beta_damage <- - 20000`

to `beta_damage <- 0`,

as in the next line that is commented out. Run the entire script again and read the output.

d) Copy and paste the new regression model estimates after the commands `summary(lm_full_model)` and `summary(lm_no_damage)`.

e) For this new set of regressions, compare the estimated coefficient for $ACCIDENT_i$ with and without the damage variable. How does this relate to the new coefficient for $DAMAGE_i$?

f) Compare the values of R^2 (labeled `Multiple R-squared`) and $\bar{R}^2$ (labeled `Adjusted R-squared`) for the two models. Now which model do you recommend (again, pretending that you don't know the true model)?