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

Assignment 4

Due Sunday, November 24, 2024 at 11:59 PM
in digital form on Webcourses.

Question 1:

Continue the in-class exercise using the script `OLS_on_repeat_A4.R` in RStudio, which is a modified version of the script `OLS_on_repeat_A4.R`. In this exercise, we repeatedly generate simulated data for car prices, 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 observe the output from the simulation. In particular, observe the statistics printed at the bottom.

a) Copy and paste the means and standard deviations from the output after the commands `sapply(reg_results[, full_list_of_variables], mean)`

and `sapply(reg_results[, full_list_of_variables], sd)`.

b) For each parameter, calculate the distance between the mean estimate and the true value, in terms of the number of standard deviations of that variable. The true values of the coefficients are listed in lines 75 to 78 in the script `OLS_on_repeat_A4.R`.

c) Now compare the average values of each of the estimated coefficients with their true values. Are they biased or unbiased? Keep in mind that with an unbiased estimator a difference of less than 2 standard deviations could often happen by chance.

Sometimes the mileage on a car is inaccurately reported: the speedometer might have broken, been replaced, or the seller may have rolled back the mileage. Now consider this situation in which the true `mileage` variable is unobserved but another measurement `mileage_1` is observed. In other words, mileage is measured with error. To implement this in the script `OLS_on_repeat_A4.R`, change part of line 146 from

```
list_of_variables <- c('mileage', 'accident', 'damage')
```

to

```
list_of_variables <- c('mileage_1', 'accident', 'damage')
```

Run the entire script again and observe the new output.

d) Copy and paste the means and standard deviations from the output after the commands `sapply(reg_results[, full_list_of_variables], mean)`

and `sapply(reg_results[, full_list_of_variables], sd)`.

e) For each parameter, calculate the distance between the mean estimate and the true value, in terms of the number of standard deviations of that variable.

f) Now compare the average values of each of the estimates with their true values when the true `mileage` is unobserved but inaccurate `mileage_1` is observed instead. Are they biased or unbiased? Again, keep in mind that with an unbiased estimator a difference of less than 2 standard deviations could often happen by chance.