Econometrics

Assignment 02

Sumit Mishra

Krea University | MSL7170 (IIT Jodhpur)

Instructions and Advice

GRADING Maximum points: 40. This assignment accounts for 20 percent of your final grade.

SOFT DEADLINE Thursday, 2021-04-15 (8pm)

HARD DEADLINE Friday, 2021-04-16 (8am)

OBJECTIVE(s) The objective of this assignment is to test your understanding of the linear regression model.

ADVICE Please be brief and precise in your answers. Unnecessarily lengthy answers will attract penalty.

WARNING #1 Code and answers copied from the internet sources will get you zero on the entire assignment.

WARNING #2 Please do not include any absolute folder path in your code. Save all downloaded datafiles to data subfolder of your course folder.

Submission Instruction #1 You must submit your R script and a PDF document with all your answers, in the following format `A02_RollNumber.R` and `A02_RollNumber.pdf` where RollNumber will be replaced by your own IIT Jodhpur ID (for instance, P20MS001).

Submission Instruction #2 You must email me your code (in an R file) and your answers (in a PDF file) with the following subject line: **MSL7170: Assignment 02 submission**.

Make sure that you have loaded the packages as shown below. If you need additional packages for your work, please add them in comma-separated manner in the same line of the code.

```
pacman::p_load(tidyverse, data.table, readxl, wooldridge)
```

Question 1 (3 points)

Describe the homoskedasticity assumption in plain English.

Question 2 (10 points)

A researcher is working on the relationship between fertility rate (x) and economic growth (y) for states in India. She runs the regression, and reports that the beta hat is 0.9, and the confidence interval is $[-0.08, 1.88]$. Interpret this confidence interval in simple language, and comment on the hypothesis test results based on these numbers. Make suitable assumptions wherever necessary.

Question 3 (10 + 5 points)

Suppose that you are interested in understanding the relationship between box-office performance (y) and the number of torrent downloads of movies (x). You collect data from IMDB about 200 top movies for the year, and corresponding downloads information from all major torrent websites. You find that there is a positive relationship between torrent downloads and box-office performance! Do we conclude that torrent downloads boost box-office performance? Defend your answer. If you believe that the estimate is biased, briefly explain whether there is upward or downward bias.

Question 4 (12 points)

a. Read the dataset in R and save it as an object named `regDta`. (**1 point**)

The variable description is given below.

Variable Name	Description
DistrictID	District ID
StateID	State ID
area11	Name of the District
pNoAssets	Proportion of households without any assets
pSC	Proportion of SC population
pST	Proportion of ST population
pSCST	Proportion of SC-ST population
pLit	Proportion literate
pop	Total population
sexratio	Sex ratio
pAgri	Proportion workforce engaged in agriculture

b. Estimate the relationship between `pSCST` (x) and `pNoAssets` (y) using simple linear regression model using the `lm( )` function in R, and write the model equation. Interpret the coefficient in plain language. (**1 + 1 points**)

c. Estimate now the relationship between `pSC` (x_1), `pST` (x_2), and `pNoAssets` (y) using the `lm( )` function in R, and write the model equation. Interpret the coefficients in plain language. (**1 + 1 points**)

d. Put together the results from the previous parts (b and c) into a clean, formatted table. (**2 points**)

e. Using part b, compute the standard error for `pSC` (x_1) manually in R. Recall that

$$se(\hat{\beta}) = \frac{\hat{\sigma}}{\sqrt{(n)} \times \sqrt{SST_j} \times \sqrt{(1 - R_j^2)}}$$

Make sure that you write your answer in your document as well. (**3 points**)

f. Our goal is to predict `pNoAssets` for Indian districts. Sketch a model- you need to list a set of explanatory variables- that will minimise omitted variable bias. You don't have to run anything in R, just have to describe the model. (**2 points**)