BT3102 Assignment 1

Jan 19 to Feb 9, 2021

Q1.

I. Which of the 4 data violate(s) the rank condition in OLS regression?

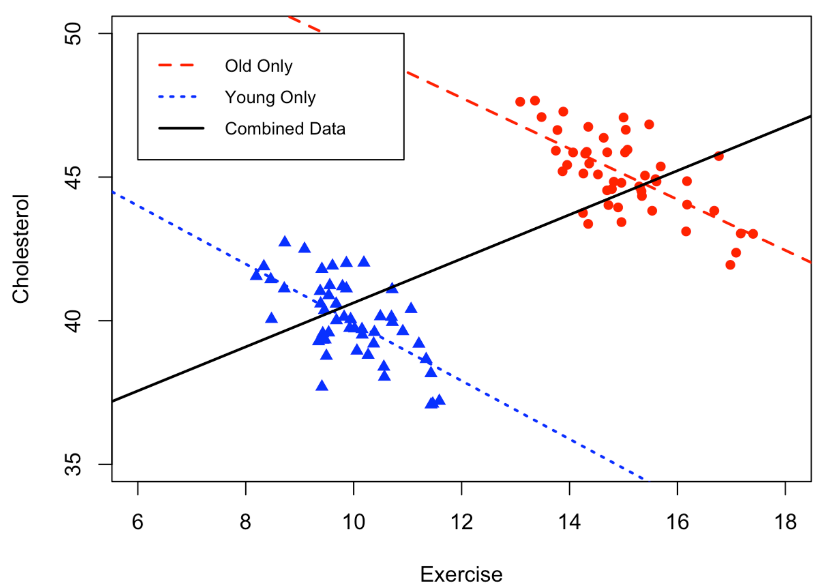
$$A.\begin{bmatrix}1 & 1\\2 & 2\end{bmatrix}\quad B.\begin{bmatrix}1 & 1\\2 & 2\\1 & 2\end{bmatrix}\quad C.\begin{bmatrix}1 & 1\\2 & 0\end{bmatrix}\quad D.\begin{bmatrix}1 & 2 & 3\\2 & 0 & 2\end{bmatrix}$$

II. Suppose that you wanted to build a model of the approval ratings of major party nominees for US president. You included the following 4 independent variables:
Years holding elected office;
Party;
Gender;
Indicator variable for being married to a former president.
Does this violate the rank condition? Explain.

III. Give a different real-world example where the rank condition fails.

Q2. Suppose you want to investigate the relation between exercise and cholesterol. Download the data "exercise_and_cholesterol.csv".

I. Reproduce the following graph. The red line is the OLS fitted line using only data for the old people. The blue line is the OLS fitted line using only data from the young people. The black line is the OLS fitted line using all the data.



II. Are the three lines giving consistent or conflicting insights? Explain why.

III. Can you propose an alternative estimation using all the data in order to obtain the same insight as the two separate regressions using only the old or the young people?

Q3. Consider the model y=0.7+2x+ϵ, where x and ϵ follow independent standard normal distributions.

I. Generate a random sample with 1000 observations. Using the simulated data and the following four computational methods to estimate the value of the model parameters. Do the four methods give the same estimates?
   a. Use R native function lm().
   b. Use the matrix algebra $\hat{\beta} = (X^T X)^{-1} X^T Y$.
   c. Recall that OLS estimate minimize the sum of squared errors. Formally $\hat{\beta} = \mathrm{argmin}_{\beta} \sum_{i=1,..,N} (y_i - \beta_0 - \beta_1 \times x_i)^2$. Use R native function optim() to find the optimizer.
   d. Use the results in Exercise 1-1 on page 19 in BT3102_OLS.pdf.

II. Now generate S =10000 independent random samples each with N observations. For each of the random samples, run an OLS regression (You can use any of the methods in I). Record all the S estimates on x (i.e., $\hat{\beta}_1$).
   a. Plot the distributions of $\hat{\beta}_1$ when N=25, 100. Compare the means and variances of the distributions. Are they different? Do they follow normal distributions?
   b. If ϵ follows (A) uniform distribution (i.e., ϵ=runif(N)) or (B) Cauchy distribution (i.e., ϵ=rcauchy(N)), do your answers in II-a change?

Q4. A car loan of $10000 was repaid in 60 monthly payments of $250, starting one month after the loan was made. Find the monthly interest rate r (error<10^{-8}).

I. Use Bisection method (code by yourself).
II. Use Newton method (code by yourself).
III. Plot the search histories for I and II using different starting points. Compare the speed of convergence.

Q5. Your market report predicts that market condition x is distributed on [0,1] with probability distribution function $f(x) = 3x^2$. When market condition is x, the return of investing y billion dollars is $p\,(x, y) = y \times (\log(x) + 1) - y^2\sqrt{1 - x^4}$.

I. Use inversion sampling to simulate 10^6 random draws for market condition x.
II. Calculate the expected returns using your draws. Plot the expected returns for $y \in (0,1)$.
III. Compute the optimal investment y^*.