MAT 5030: Homework 4

- Assigned on October 26 (Wed), 2016.
- **Due at 5:30pm on November 2 (Wed)**, 2016.
- Submit **by e-mail at seikibunpu@gmail.com**
- The e-mail title will be "M5030 HW4 - Your Name".
- Use the following format.
  - R-code, output, and brief explanations in one TXT or MS-Word file. Cut and paste as text. Do not use screenshots.
  - Graphics in **ONE** MS-Word or PDF file.
  - In case of MS-Word, you can put both in one file.
- Truncate the output if a single output exceeds 20 lines in the R console.
- A late submission will reduce your score as follows. E.g, if you submit 19 hours late and your work gets 9 points out of 10, your score will be $9 \times 0.8 = 7.2$ out of 10.

Delayed by (in days)	0-0.5	0.5-1	1-1.5	$\cdots$	4-4.5	4.5-
Multiplier	0.9	0.8	0.7	$\cdots$	0.1	0

1. A researcher surveyed household income (in thousand dollars) and Engle's coefficient (percentage of income used for food) of nine housesholds in the city A, and the result is as follows:

household	A	B	C	D	E	F	G
Income	35	90	47	45	68	70	55
Engel's coef.	21	8	16	14	11	12	10

   (a) Make a scatter plot (X: Income, Y: Engel's coefficient).

   (b) Regress Engel's coefficient Y on Income X with a linear model $Y = a + bX + \epsilon$ by `'lm'` function.

   (c) Do the same estimation as (b) by matrix algebra in R (without using `'lm'` function).

   (d) Do the same estimation as (b) by numerically minimizing sum of squared errors. (Hint: The 'optim' function finds minimizer of a function. See Exercise 8 for details.)

2. Do Exercise 2 in the Chapter 6 slides.

3. The attached "GDPLiteracy.csv" file includes GDP per capita (in 2009) and literacy rate (in 2009 or latest) of 143 countries and regions (source: Social Indicators, United Nations at `http://unstats.un.org/unsd/demographic/products/socind/`).

(a) Make a scatter plot for literacy rate (x-axis) and GDP per capita (y-axis). What problems do you see to apply simple linear regression?

(b) Transform GDP per capita by natural logarithm, and make a scatter plot again.

(c) Regress log of GDP per capita on literacy rate. Report intercept, slope, and their standard deviations and t-statistics. Is the slope significantly different from zero?

(d) Overlay a prediction band on the scatter plot in (b).

(e) Make a residual plot. Do you see any patterns?

(f) Make a normal Q-Q plot for residuals. Residuals are approximately normal?

(g) If we can add one more independent variable to predict GDP per capita, what variable will you add? (You do not have to do any analysis for (g)).