Week #10 - Assignment

Multiple Linear Regression

Due: 04/14/2021

Family caregiving of older adults is more common in Korea than in the United States. A research team studied 100 caregivers of older adults with dementia in Seoul, South Korea. The dependent variable was caregiver burden as measured by the Korean Burden Inventory (KBI). Scores ranged from 28 to 140, with higher scores indicating a higher burden. Explanatory variables were indexes that measured the following:

ADL: total activities of daily living (low scores indicate that the elderly perform activities independently).

MEM: memory and behavioral problems (higher scores indicate more problems).

COG: cognitive impairment (lower scores indicate a greater degree of cognitive impairment).

The reported data are given in the following code chunk.

```
Y = c(28, 68, 59, 91, 70, 38, 46, 57, 89, 48, 74, 78, 43, 76, 72, 61, 63, 77, 85, 31, 79,
      92, 76, 91, 78, 103, 99, 73, 88, 64, 52, 71, 41, 85, 52, 68, 57, 84, 91, 83, 73, 57,
      69, 81, 71, 91, 48, 94, 57, 49, 88, 54, 73, 87, 47, 60, 65, 57, 85, 28, 40, 87, 80,
      49, 57, 32, 52, 42, 49, 63, 89, 67, 43, 47, 70, 99, 53, 78, 112, 52, 68, 63, 49, 42,
      56, 46, 72, 95, 57, 88, 81, 104, 88, 115, 66, 92, 97, 69, 112, 88)

X1 =c(39, 52, 89, 57, 28, 34, 42, 52, 88, 90, 38, 83, 30, 45, 47, 90, 63, 34, 76, 26,68,
      85, 22, 82, 80, 80, 81, 30, 27, 72, 46, 63, 45, 77, 42, 60, 33, 49, 89, 72, 45, 73,
      58, 33, 34, 90, 48, 47, 32, 63, 76, 79, 48, 90, 55, 83, 50, 44, 79, 24, 40, 35, 55,
      45, 46, 37, 47, 28, 61, 35, 68, 80, 43, 53, 60, 63, 28, 35, 37, 82, 88, 52, 30, 69,
      52, 59, 53, 65, 90, 88, 66, 60, 48, 82, 88, 63, 79, 71, 66, 81)

X2 =c(4, 33, 17, 31, 35, 3, 16, 6, 41, 24, 22, 41, 9, 33, 36, 17, 14, 35, 33, 13, 34, 28,
      12, 57, 51, 20, 20, 7, 27, 9, 15, 52, 26, 57, 10, 34, 14, 30, 64, 31, 24, 13, 16,
      17, 13, 42, 7, 17, 13, 32, 50, 44, 57, 33, 11, 24, 21, 31, 30, 5, 20, 15, 9, 28,
      19, 4, 29, 23, 8, 31, 65, 29, 8, 14, 30, 22, 9, 18, 33, 25, 16, 15, 16, 49, 17, 38,
      22, 56, 12, 42, 12, 21, 14, 41, 24, 49, 34, 38, 48, 66)

X3 =c(18, 9, 3, 7, 19, 25, 17, 26, 13, 3, 13, 11, 24, 14, 18, 0, 16, 22, 23, 18, 26, 10,
      16, 3, 3, 18, 1, 17, 27, 0, 22, 13, 18, 0, 19, 11, 14, 15, 0, 3, 19, 3, 15, 21, 18,
      6, 23, 18, 15, 15, 5, 11, 9, 6, 20, 11, 25, 18, 20, 22, 17, 27, 21, 17, 17, 21, 3,
      21, 7, 26, 6, 10, 13, 18, 16, 18, 27, 14, 17, 13, 0, 0, 18, 12, 20, 17, 21, 2, 0,
      6, 23, 7, 13, 13, 14, 5, 3, 17, 13, 1)
```

In this assignment, you are expected to replicate the analysis in case study #1 in the weekly note. To be more specific, you can proceed with the analysis with the following steps.

1. Perform exploratory data analysis: pair-wise scatter plot. Describe the patterns you observed from the pair-wise plot.
2. Build an initial model that includes all variables and perform the residual analysis. Inspect the residual plot and describe potential abnormal patterns you observed from the residual plots.
3. Explore potential power transformation using Box-cox transformation. Use trial-and-error to identify an ppropriate range so you can view the 95% confidence interval for λ. Please keep in mind that $\lambda = 0$ implies log-transformation.
4. Use the automatic variable selection to identify the **best** model.
5. Use the coefficient of determination to select the final model from the candidate models (initial, the **best** model from step 3.)
6. Interpret the final working model.