

Stat123 – Regression model

Using data `stat123_regression.csv`, perform the following analyses and answer the questions. Complete this project using R. RMarkdown is recommended as well.

Description of dataset:

*Son et al. (2003) studied 100 caregivers of older adults with dementia in Seoul, South Korea. The dependent variable was caregiver burden measured by the Korean Burden Inventory (KBI). Scores ranged from 28 to 140 with higher scores indicating higher burden. Explanatory variables were indexes that measured the following:

ADL: (X_1) total activities of daily living (low scores indicate that the elderly perform activities independently).

MEM: (X_2) memory and behavioral problems (higher scores indicate more problems).

COG: (X_3) cognitive impairment (lower scores indicate a greater degree of cognitive impairment).

1. After reading the data into R using the `read.csv` function, provide summary of the data. Comment on your results and especially on any unusual features in the data.

```
regression.dat <- read.csv('stat123_regression.csv')
```

2. Produce the boxplot of ADL. Comment. Is there any outliers?

3. Produce the scatterplots of Y and the X 's. Comment. Is a linear model appropriate for this data? Why or why not? Are the X 's correlated amongst themselves.

4. Fit univariable linear models, Y versus X_i for each of the three X regressor variables.

- What are the estimated regression models?
- Compare the univariable models above.
- Check the fit of the models and comment.
- Explain each of the estimated regression parameters (except the intercept) in words.

*Son, G, Wykle, M. and Zauszniewski, J. "Korean Adult Childre Caregivers of Older Adults with Dementia," *Journal of Gerontological Nursing*, 29 (2003), 19-28.