# **BS806 --- Homework 5**

**Reading**. Chapters 9 of (SW) Sanford Weisberg. Applied Linear Regression, 4th ed. Wiley

**Question 1 (100 points)**. Use the data file ***pulsedata.csv*** for this assignment

Between 1995 and 1998, students in an introductory statistics class taught by Professor John Eccleston and Dr. Richard Wilson at The University of Queensland in Australia took part in a simple exercise experiment. First, the students were randomly assigned to either a sitting group or a running group. Then all of the students took his/her own pulse. Then, depending on the pre-assigned group, each student either sat for a minute or ran in place for a minute. After that, each student took his/her pulse again.

We are interested in determining which variables explain the largest proportion in variability in the student’s first pulse measurement.

|  |  |
| --- | --- |
| **Variable** | **Description** |
| Height | Height (cm) |
| Weight | Weight (kg) |
| Age | Age (years) |
| Sex | Sex (1 = male, 2 = female) |
| Smokes | Regular smoker? (1 = yes, 2 = no) |
| Alcohol | Regularly consumes alcohol? (1 = yes, 2 = no) |
| Exercise | Frequency of exercise (1 = high, 2 = moderate, 3 = low) |
| Ran | Did the student run or sit between the pulse measurements? (1 = ran, 2 = sat) |
| Pulse1 | First pulse measurement (beats per minute) |
| Pulse2 | Second pulse measurement (beats per minute) |
| Year | Year of class (95 – 98) |

**Model fitting**

1. (10 points) Run a multiple linear regression with *pulse1* as the outcome variable and *height*, *weight*, *age*, *sex*, *smokes*, *alcohol*, *exercise* as the predictor variables and Conduct a hypothesis test to examine whether pulse1 is linearly associated with *height*, *weight*, *age*, *sex*, *smokes*, *alcohol*, and *exercise*. Write a full report.

**Model Diagnostic**

1. (20 points) Check normality assumption. Explain your reasoning and provide formal statistical test and visual evidence. (For simplicity, we will assume that all of the other regression assumptions are met.)
2. (20 points) Check the outliers. Conduct the analysis to examine whether there are any potential outliers. If so, please list their ID. Explain your reasoning and provide the evidence for your analysis.
3. (20 points) Check for influential points. Conduct the analysis to examine whether there are any potential influential points. If so, please list their ID. Explain your reasoning and provide the evidence for your analysis.
4. (20 points) Check for leverage points. Conduct the analysis to examine whether there are any potential leverage points. If so, please list their ID. Explain your reasoning and provide the evidence for your analysis.

**Decision**

1. (10 points) Based on your analysis above, is there any data points concerning you? If so, what’s your suggestion for the next step.