**LINEAR REGRESSION**: Homework 

*Professor Jingchen Liu*

Fan Yang

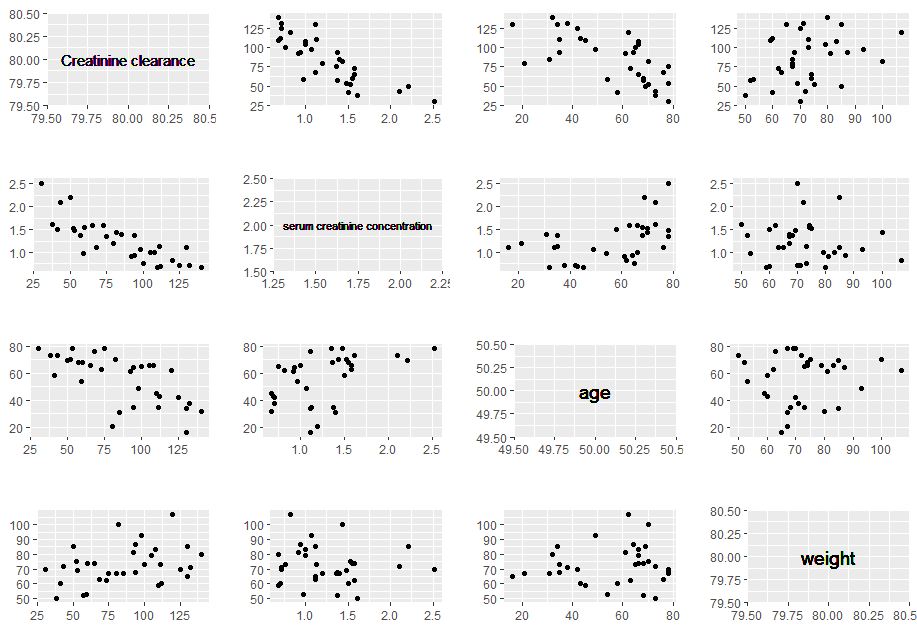
UNI: fy2232

# Problem 1 (9.9)



# Problem 2 (9.15)

## (b)





The response variable and the variable serum creatinine concentration () show significant linear relationship. And with the variable age () also forms a linear line but it is not as linear as with . As for the last variable weight (), points are uniformly lied on the plot with little evidence of linearity.

In the correction matrix,  and  have a correction of 0.468, which may imply multicollinearity problem.

## (c)



According to the result above, F-statistic is large with p-value less than 0.01. And each variable coefficient shows significant contribution to the model. So, all the variables should be attained.

# Problem 3 (9.16)

## (a)

First compute a function to calculate the Cp criteria:



Then we list all the 511 kinds of combination of the 9 variables and store the combinations in a list called *col*.

Finally calculate all 511 Cp of the 511 combinations and list the three lowest value.



 represent variable combinations

 respectively.

and correspond to column 2,3,4,5,6,7,8,9 respectively.

As we can see the three lowest value are 

and the three corresponding variables combination is



## (b)

The three lowest value are , so there is little difference between the three subset models.

# Problem 4 (9.19)

## (a)



Using the stepwise function and get the best subset of variables. The result subset is 1,2,3,4 which corresponds to 

so the model becomes 