**LINEAR REGRESSION**: Homework 

*Professor Jingchen Liu*

Fan Yang

UNI: fy2232

# Problem 1 (3.14)

## (a)

Hypothesis:



if , then conclude 

if , then conclude 

Let’s compute the lack of fit test:



The lack of fit statistic is  with p-value , supporting the linearity of the regression model.

## (b)

Having an equal number of replications at each of1he X levels would lead to a smaller error term and get a better fit regression model. But all the effects covered in the error term could not vary at random from one repeated observation to the next because all the Y in the same group are set the same.

## (c)

When it leads to nonlinear conclusion, the test in part (a) cannot indicate what regression function is appropriate. I will try to make box-cox transformation on the data and conduct regression on the transformed data.

# Problem 2 (7.7)

## (a)





## (b)

Hypothesis:



if , then conclude 

if , then conclude 





Because , with p-value , then we conclude .

# Problem 3 (7.10)

Hypothesis:



if , then conclude 

if , then conclude 



Because , with p-value , then we conclude .

# Problem 4 (7.16)

## (a)



## (b)

As computed in part (a), 

## (c)



which is the same result as Problem 6.5b.

# Problem 5 (7.24)

## (a)



The fitted regression function is 

## (b)

The two regression coefficients for moisture content are both 4.425.

## (c)



Therefore, 

## (d)

In the correlation matrix obtained in Problem 6.5a, the correlation between  and  is 0, which corresponds to the results in parts (b) and (c). The two variables contribute independently to .

# Problem 6 (7.37)

## (a)

## (b)

The variable  number of hospital beds has the largest coefficient of partial determination. When  is added to the model, the error sum of squares is reduced 6.17%, which is the most among the four new variables.



The extra sum of squares of  is the largest among the 4 variables.

## (c)

Hypothesis:



if , then conclude 

if , then conclude 





Because , with a very small p-value, then we reject , which means the variable  number of hospital beds is helpful in the regression model.

The coefficient of partial determination of the other 3 variables are much smaller than that of , so the F\* test statistics for the other three potential predictor variables will not be as large as the one here.

Actually, the F\* test statistics for  are 12.94069, 1.681734, 3.216562, respectively, which is much smaller than that of  541.1801.