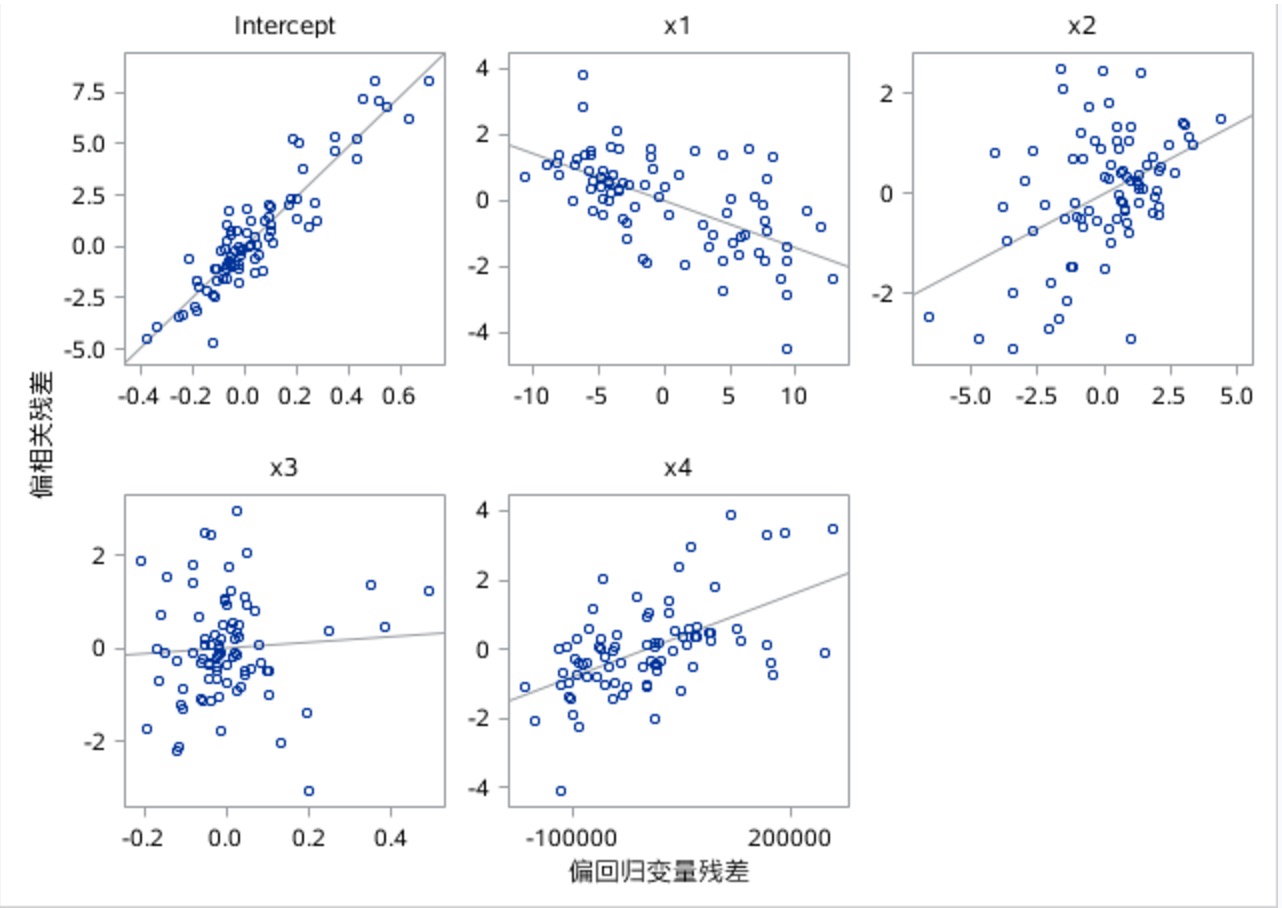
**HW 10**

**Fangling Zhang**

**Q8**

**(a)**



**(b)**

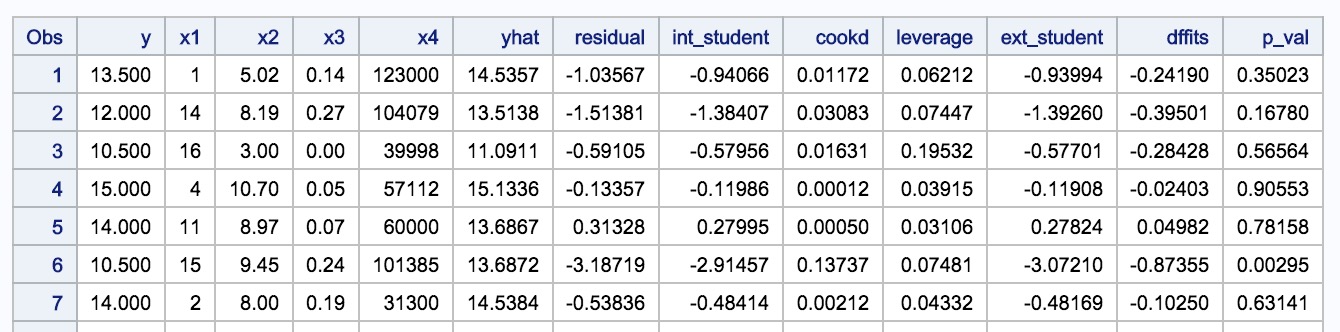
AS the curvilinear relation between y and x1,x2, x4 are strong positive or negative, these tree variables are useful for the fitted regression function.

However, the plot of x3, vacancy rates is not helpful for the regression model because it shows a horizontal band. Thus it does not "add value". It doesn’t give additional information for the fitted regression function.

**Q12**

**(a)**

From the output of SAS, the studentized deleted residuals are: res(6)=-3.07210, res(62)=2.66808.



The Bonferroni outlier test:

Decision rule:

p-value<0.01=, reject H0,conclude xi is an outlinear; p-value>0.01= do not reject H0, conclude xi is not an outlinear.

Conclusion:

The p-value of test 6 is 0.0029<0.01, reject H0, case 6 is an outliner;

The p-value of test 62 is 0.0068<0.01, reject H0, case 62 is an outlier.

(b)











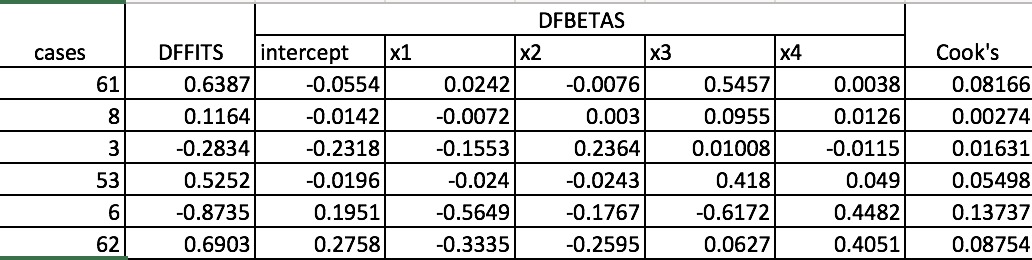
From the output of SAS, the h matrix are: h(3)=0.1953, h(8)=0.2022, h(53)=0.1792, h(61)=0.3037, h(65)=0.1291. Here 2p/n=2\*5/81=0.12346.

Therefore, the h matrix of case 3, 8, 53 and 61 are larger than 2p/n and these four cases are outlying X observations.

(c)

, it is well within the range of leverage values hii for the cases in the data set, no extrapolation is involved.

**(d)**

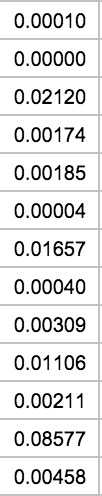
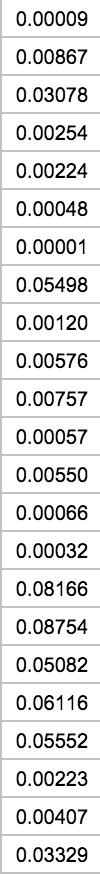
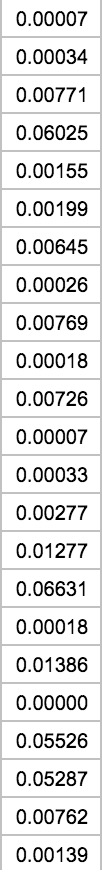
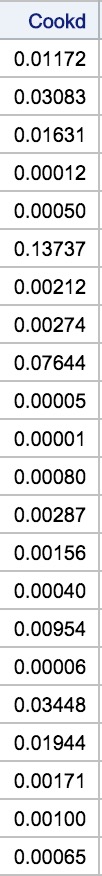
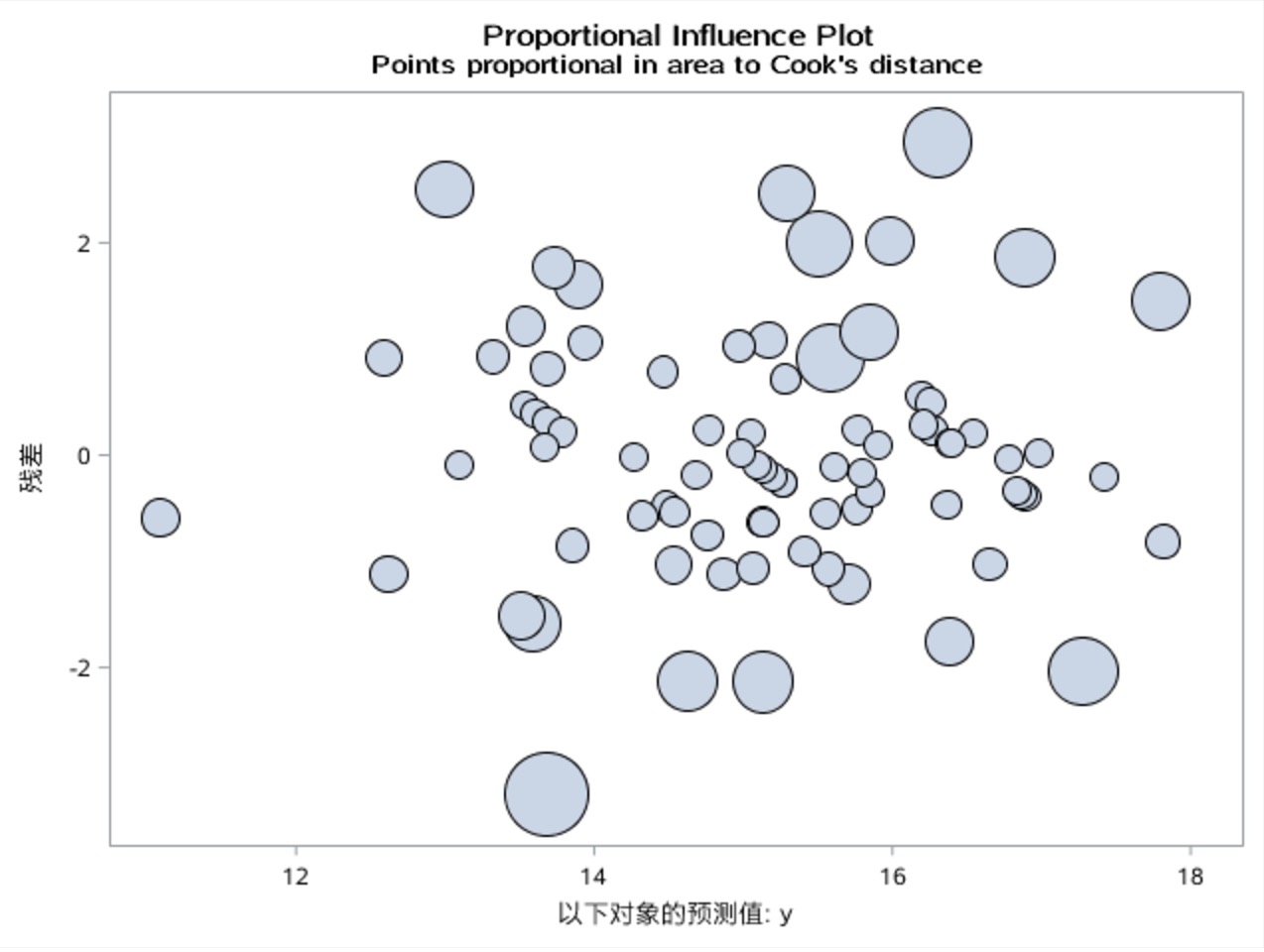


The absolute value of DFFITS are all smaller than 1, so the 6 cases are not influential;

DFBETAS: 2/ the b2 of x3 >0.2222, it is larger a lit bit, so it is not influential; All the coefficients of case 53 are larger than 0.2222, it is influential. Case 6 is influential; Case 61 and 62 are exceeds a little bit.

COOK': The D(3)=0.01631, D(6)=0.13737, D(8)=0.00274, D(53)=0.05498, D(62)=0.08754, D(61)=0.08166. Considering Cook’s distance, values for all cases are quite small, so they are all influential.

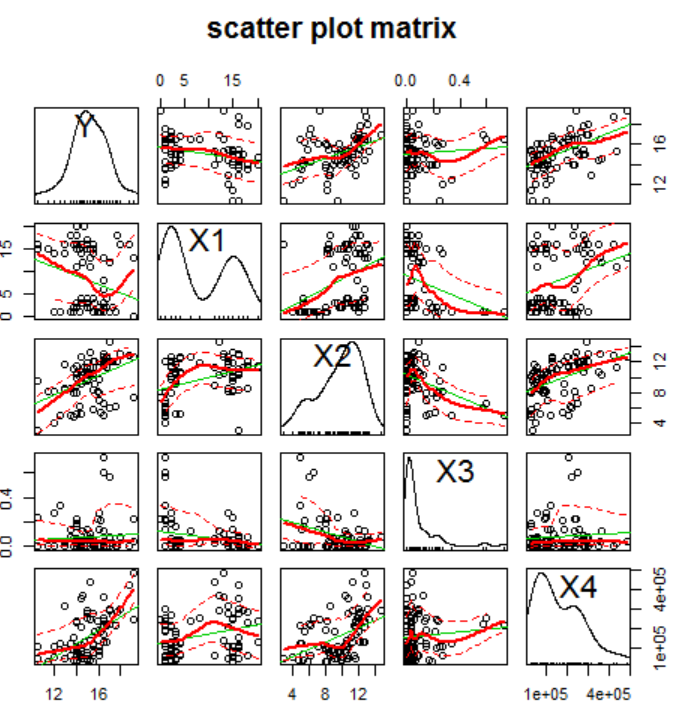
**(f)** All cook’s distance are as shown.

From this graph, f(p, n-p)= F(5, 76)=3.16> 3.14，all the cook’s distance are smaller than it. It means all cases are influential.

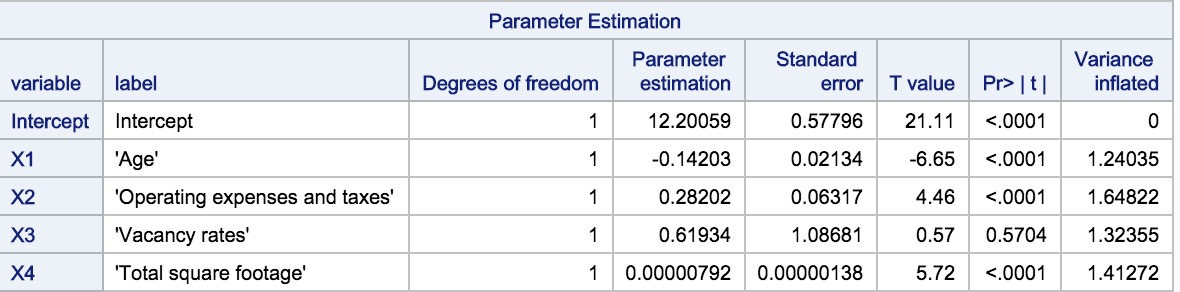
**Q18**

(a)



From the above figure, we can see that Y, X2 and X4 have a positive correlation with each other. x1 x2 x4 have a positive correlation with each other. x1 and x3 has a negative correlation.

(b)



the (VIF)1= 1.24035, (VIF)2=1.64822, (VIF)3=1.32355, (VIF)4=1.41272; they are all larger than 1. So there are multicollinearity problems in variables.

**Q26.**

`