**HW 11**

**Fangling Zhang**

Q2.



Robust regression just decrease the undue influence of outlying cases, but not wipe out their influence. Therefore it is not true that we do not need to worry about outliers and influential observations at all.

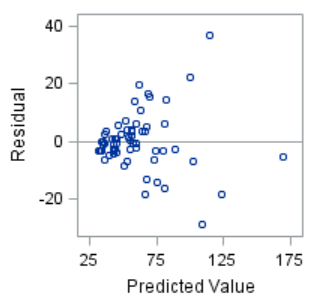
Q8.

1. Create two indicator variables for highest degree attained:



(b)

With a first-order model and ordinary least squares, the estimated parameter and plot of residuals against predicted y is as follows. The residual plot suggest that the error variance increase obviously as predicted y increase. In this model, the error variance is not constant.

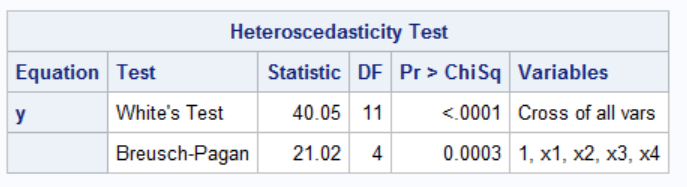


(c)

**The alternatives:** H0: Constant variance; Ha: Heteroskedasticity

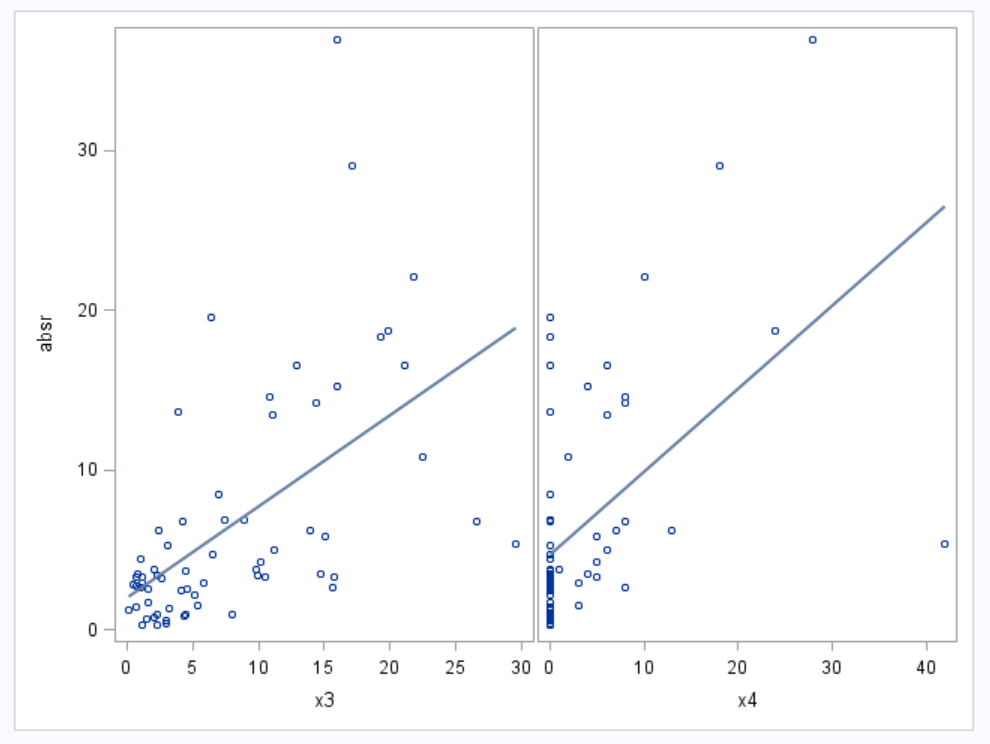
**The decision rule:** If p-value above an appropriate threshold (e.g. p>=0.05), conclude H0; If p<0.05, conclude Ha;

With SAS, both result of the White and Breusch-Pagan reject the null hypothesis of no heteroscedasticity, and we conclude that there exists heteroscedasticity. This implies that the standard errors of the parameter estimates are incorrect and, thus, any inferences derived from them may be misleading.



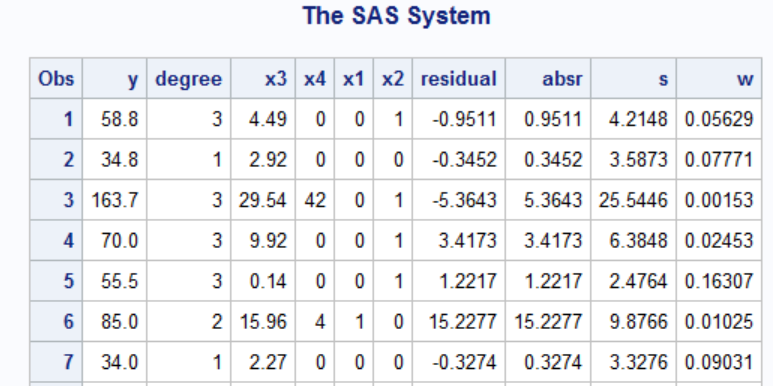
(d)

These absolute residuals plots suggest that residuals variance increase with both X3 and X4.



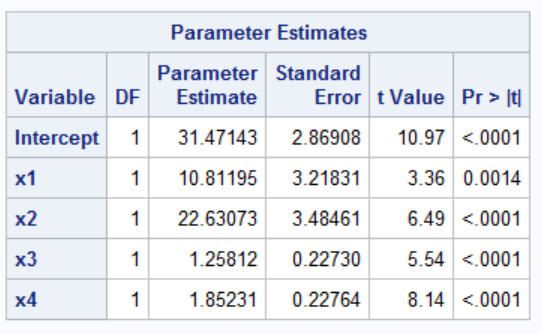
(e)

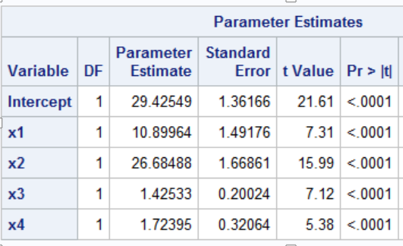
Use a linear regression of absolute residual on both X3 and X4 to estimate the weights;



(f)

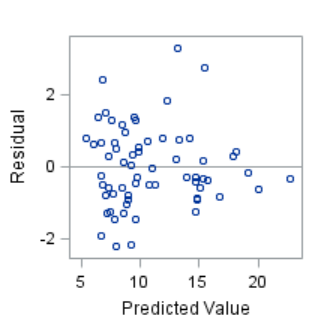
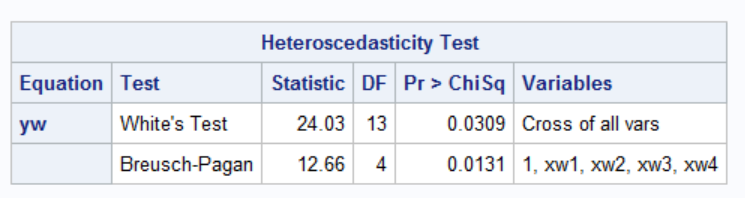
The weighted least squares estimates of the regression coefficients are close, but not equal, to the ones obtained with ordinary least squares.

The weighted least squares estimates of parameters vs The ordinary least squares estimates of parameters



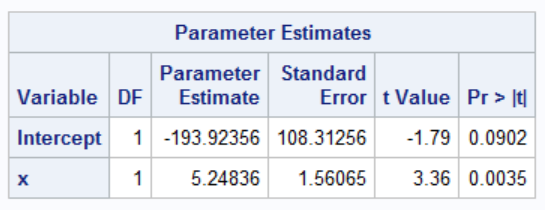
(g) The standard error values of weighted estimates are generally much smaller than those of ordinary estimates.

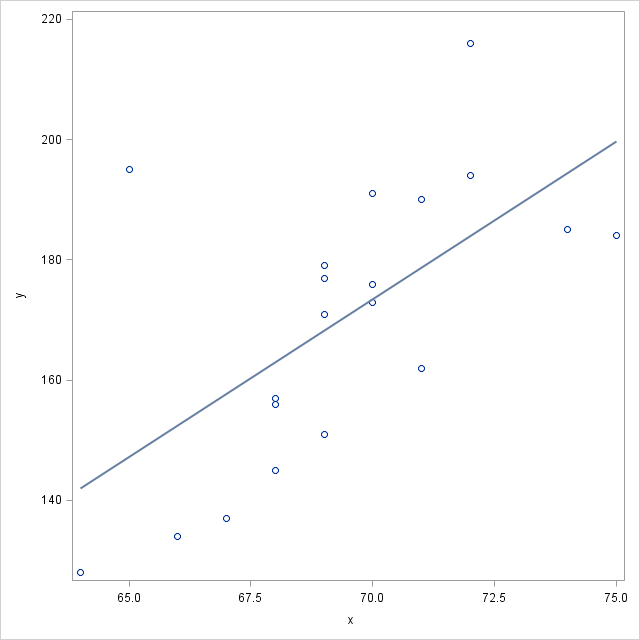
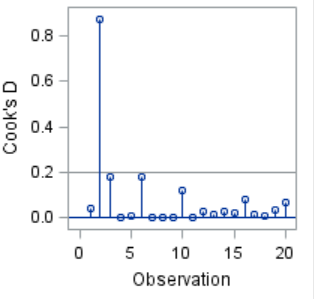
(h) From the OLS fit, the residuals plot versus fitted values and the result of White and Breusch-Pagan tests are as follows. They suggest that the error variance are much more constant. There is not obvious heteroscedasticity now.



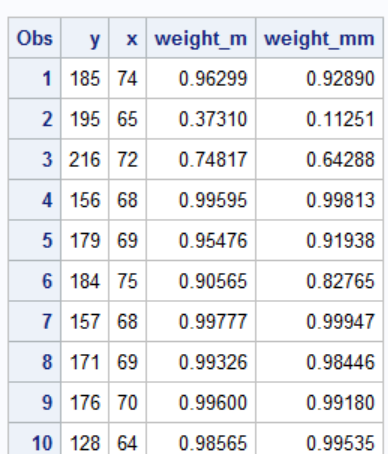
**Q12**

The scatter plot with fitted regression function suggests that this simple linear model fits these data well except some outliers. The plot of Cook’s distance suggests that there is an obviously influential outlier.





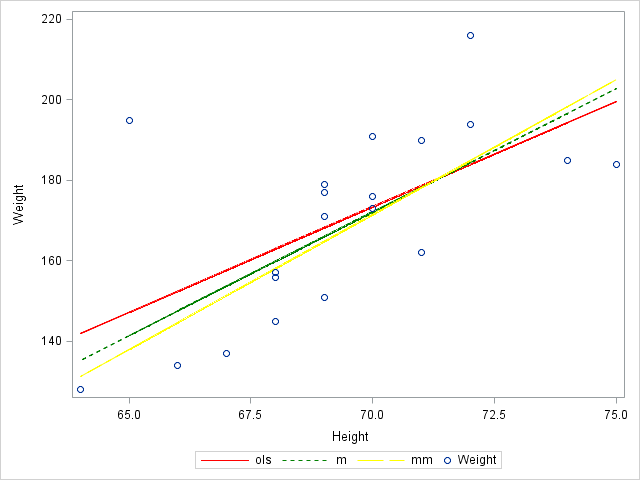
(b)



From the above figure, the weights of MM-estimate are generally smaller than those of M-estimate, except for a few cases have large M-estimate weight. That is because MM-estimate proceeds by finding a highly robust and resistant S-estimate that minimizes an M-estimate of the scale of the residuals. Thus for cases far away from regression line, the weights of MM-estimate are smaller than those of M-estimate. But for a few very close cases, the weights of MM-estimate are a little larger than those of M-estimate.

(c)

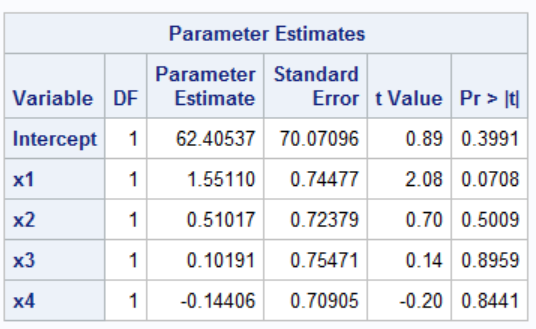
Superimpose all three fitted lines on a scatterplot of the data. There is some difference, m’ skewness is a litter larger than ols’ and mm’s is larger than m’s.

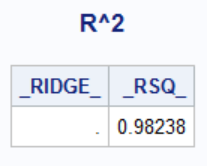


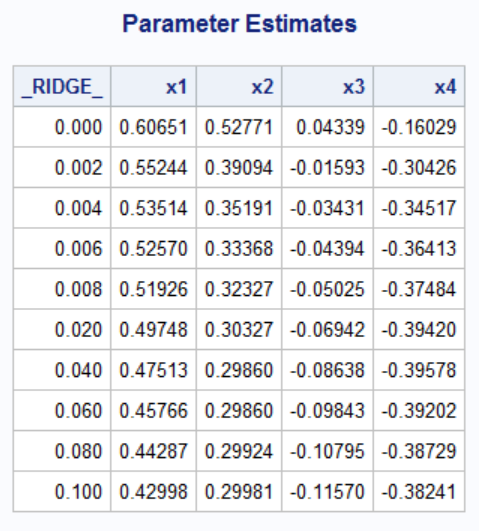
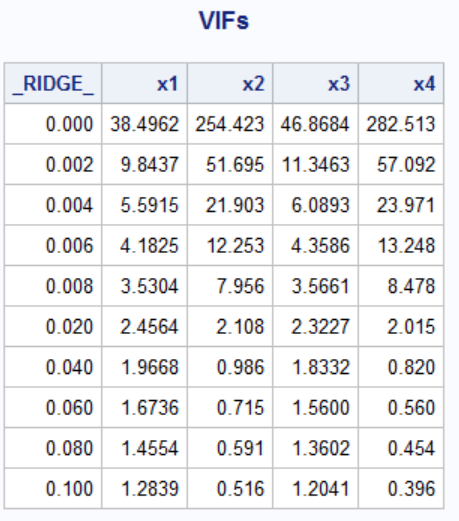
Q13

Q23

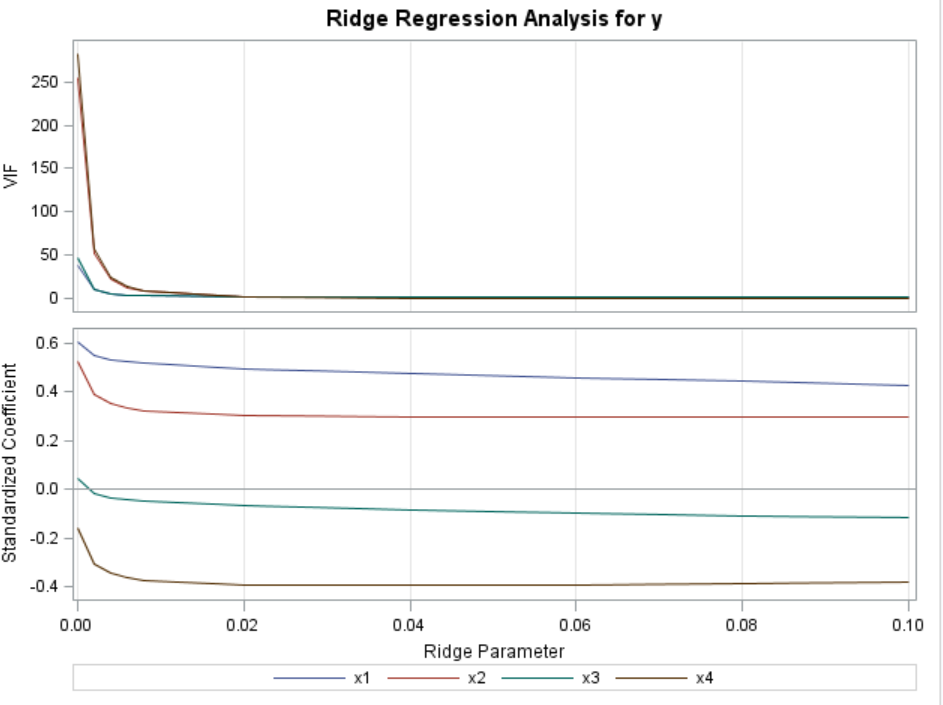
1. The estimated regression function: y=62.405+1.551x1+0.510x2+0.102x3-0.144x4.







1. Yes, the ridge regression coefficients exhibit substantial changes near c=0.



(d)

C=0.04 is proper here because for this value, VIF near 1 and the estimating regression coefficients appear to have become reasonably stable.

(e)

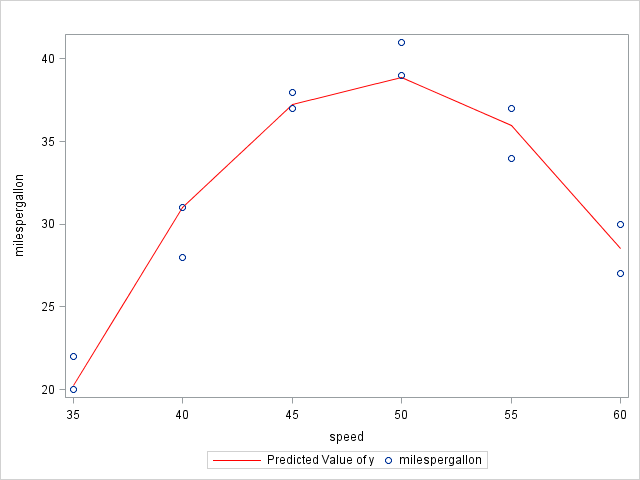
When C=0.04, the function is y\*=0.475X1\*+0.299X2\*-0.086X3\*-0.396X4\*

After transform y=82.912+1.215X1+0.288X2-0.203X3-0.356X4

Q28

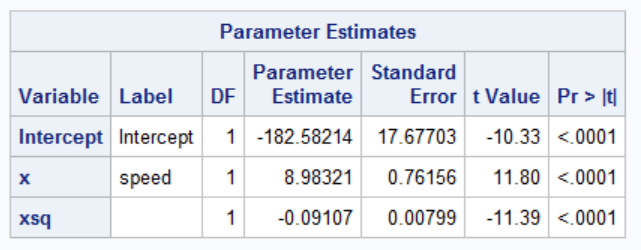
Note that in part (e), the reference is to (11.58) and (11.59), not (11.56).

(a)



The quadratic regression function appear to be a good fit here.

(b)



From the above figure, we get b1=8.983, b11=-0.091,

=96.857

The responding estimated mean mileage = -182.582 + 8.983\*96.857 - 0.091\*96.857\*96.857= -166.212.