a)

house<-read.csv(file.choose(), header=TRUE)

attach(house)

model1=lm(Value ~ Size + Age + Garage + Rooms + Baths)

rstandard=rstandard(model1)

rstandard[order(rstandard)]

4 2 5 7 1 9 3 6

-1.7087039 -0.3303597 -0.1752872 0.1115119 0.1714062 0.2128637 0.5150996 1.2715418

8

1.6711654

> hist(rstandard)

> CooksD=cooks.distance(model1)

> hist(CooksD)

> leverages=hatvalues(model1)

> par(mfrow=c(1,3))

> hist(rstandard)

> hist(leverages)

> hist(CooksD)

>



b)

> rstandard[order(rstandard)]

4 2 5 7 1 9 3 6

-1.7087039 -0.3303597 -0.1752872 0.1115119 0.1714062 0.2128637 0.5150996 1.2715418

8

1.6711654

There are no outlier points that are outside of +-2.

> leverages[order(leverages)]

2 4 1 8 9 7 3 5

0.2982571 0.3881361 0.6172772 0.6318148 0.6869392 0.7282425 0.8063996 0.8578796

6

0.9850540

> #Cutoff = 3(5+1)/9 = 2

Since the cutoff is 2, there are no high leverage points that exceed 2.

> CooksD[order(CooksD)]

m=k+1=6, n-m=3;

qf(.95, 6, 3)

8.940645

The Cutoff point for Cook’s D is 8.940645, so Observation 6 is the only observation that exceeds the cutoff.

c)

> model2=lm(Value ~ Size + Age + Garage + Rooms + Baths, subset=-c(6))

> summary(model2)

Call:

lm(formula = Value ~ Size + Age + Garage + Rooms + Baths, subset = -c(6))

Residuals:

1 2 3 4 5 7 8 9

20.964 -2.091 -12.168 -27.632 -15.901 -17.587 28.031 26.383

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 264.6093 163.2638 1.621 0.247

Size -0.2334 0.3015 -0.774 0.520

Age -8.0253 3.9093 -2.053 0.176

Garage 195.8273 160.5691 1.220 0.347

Rooms 123.6370 85.9719 1.438 0.287

Baths -38.9214 45.9141 -0.848 0.486

Residual standard error: 41.23 on 2 degrees of freedom

Multiple R-squared: 0.9812, Adjusted R-squared: 0.934

F-statistic: 20.82 on 5 and 2 DF, p-value: 0.04645

summary(model1)

Call:

lm(formula = Value ~ Size + Age + Garage + Rooms + Baths)

Residuals:

1 2 3 4 5 6 7 8 9

5.257 -13.719 11.236 -66.261 -3.276 7.706 2.882 50.270 5.904

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 110.1807 154.2452 0.714 0.5266

Size 0.2162 0.0800 2.703 0.0736 .

Age -2.1893 1.0160 -2.155 0.1201

Garage -38.4491 57.7353 -0.666 0.5531

Rooms -0.7231 33.4941 -0.022 0.9841

Baths 1.1132 45.3529 0.025 0.9820

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 49.57 on 3 degrees of freedom

Multiple R-squared: 0.9599, Adjusted R-squared: 0.8931

F-statistic: 14.36 on 5 and 3 DF, p-value: 0.02628

Garage, Rooms, and Baths improved in the p-values. R2 improved from 0.9599 to 0.9812, and the R2 adjusted did not change much. Residual error decreased, which is beneficial to the new model. Overall, this new model has got worse.

4)

> # Any vif value that exceeds 10 denotes severe multicollinearity

> # None of our variables exceed that here.

> house<-read.csv(file.choose(), header=TRUE)

> attach(house)

> summary(house)

> cor(cbind(Size, Age, Garage, Rooms, Baths))

> model1=lm(Value ~ Size + Age + Garage + Rooms + Baths)

> summary(model1)

Size Age Garage Rooms Baths

Size 1.0000000 -0.5397071 0.54600624 0.74776719 0.6365757

Age -0.5397071 1.0000000 -0.48266724 -0.36923065 -0.3271453

Garage 0.5460062 -0.4826672 1.00000000 -0.01399731 0.8004987

Rooms 0.7477672 -0.3692307 -0.01399731 1.00000000 0.2320059

Baths 0.6365757 -0.3271453 0.80049874 0.23200592 1.0000000

#The correlation between them are not much significant,except for the Size~ROOMS and Garage and Baths.

> house1=data.frame(Size, Age, Garage, Rooms, Baths)

> pairs(house1, upper.panel=NULL)



> vif(model1)

Size Age Garage Rooms Baths

7.674844 1.818328 6.630889 5.883455 3.812678

#Since there is no data above exceeds 10,thus none is a severe multicollinearity.