CASE DIAGNOSTICS IN R

Data is the name of the data set.

Model is the name of a model that has been run.

Define the estimated standard deviation:

```
> s = summary(Model)$sigma
```

Then here is the correspondence between definitions in the book and R commands for the various diagnostic measures:

Name and notation	Formula in book	Flag values	R command(s)
Residuals = e_i	1.16	No rule	`residuals(Model)`
Semistudentized residuals = e_i*	10.9	No rule	`residuals(Model)/s`
Studentized residuals = r_i	10.20	No rule	`rstandard(Model)`
Studentized deleted residuals = externally studentized residuals= t_i	10.26	$\lvert t_i \rvert > 3$ or use Bonferroni, flag $\lvert t_i \rvert > t(1 - \alpha/2n;\ n - p - 1)$	`rstudent(Model)`
Leverage = diagonal elements of the hat matrix $\mathbf{H} = h_{ii}$	See pg 398	$h_{ii} > 2p/n$ or sometimes use $3p/n$	`hatvalues(Model)`
Difference in fits = $(DFFITS)_i$	10.30	For small to medium n, flag if $\lvert DFFITS \rvert_i > 1$; For large n, flag if $> 2\sqrt{\dfrac{p}{n}}$	`dffits(Model)`
Cook's distance = D_i	10.33	Compare to $F(p, n - p)$ and flag if $>$ about $F(.20; p, n - p)$	`cooks.distance(Model)` To find F cutoff for .20, suppose df = 4, 64: `qf(.2, df1=4, df2=64)`

Example of running the model and adding some of these:

```
Model <- lm(Salesprice~Bedrooms+LotSize+SqFt.100, data=Data)
Data$residuals <- residuals(Model)
Data$rstudent <- rstudent(Model)
Data$Hii <- hatvalues(Model)
Data$cooks <- cooks.distance(Model)
s=summary(Model)$sigma
Data$EiStar = residuals(Model)/s
```

Once you have created all of the diagnostics, there are various ways to identify which cases to "flag." For instance, the following commands will flag cases with high leverage (greater than 2p/n = 2*4/68), and print all values for those cases:

```
high = Data$Hii[Data$Hii>2*4/68]
flagH = Data[match(high,Data$Hii),]
flagH
```

You could also export the data and work with it in Excel. For example to create a comma-separated file "Mydata.csv" to use in Excel:

```
write.table(Data, "C:/Users/jutts/Documents/Mydata.csv", sep =",", col.names = T, row.names = F)
```