

CASE DIAGNOSTICS IN R

Data is the name of the data set.

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

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

Name and notation	Page in book	Flag values	R command(s)
Residuals = e_i	28	No rule	<code>residuals(Model)</code>
Standardized residuals = $stdres_i$	184	No rule	<code>rstandard(Model)</code>
Studentized residuals = $studres_i$	184	Moderate: $ t_i > 2$ Extreme: $ t_i > 3$	<code>rstudent(Model)</code>
Leverage = h_i	180 for simple 182 for multiple	Moderate: $h_i > 2(k+1)/n$ Extreme: $h_i > 3(k+1)/n$	<code>hatvalues(Model)</code>
Cook's distance = D_i	186	Moderate: $D_i > 0.5$ Extreme: $D_i > 1$	<code>cooks.distance(Model)</code>

Example of running a model and adding some of these:

```
Model <- lm(Salesprice~Bedrooms+LotSize+SqFt.100, data=Data)
Data$residuals <- residuals(Model)
Data$rstudent <- rstudent(Model)
Data$Hi <- hatvalues(Model)
Data$cooks <- cooks.distance(Model)
```

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 for $k = 3$ and $n = 68$ (flag if greater than $2(k+1)/n = 2*4/68$), 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 copy the file to the clipboard, where you can then paste it into a .txt file:

```
write.table(Data, file = "clipboard", sep = ",", col.names = T, row.names = F)
```