Influential Points: R lab(Part II)

```
?LifeCycleSavings

names(LifeCycleSavings)

rownames(LifeCycleSavings)

nrow(LifeCycleSavings)

fit=lm(sr~., data= LifeCycleSavings)


influence.measures(fit)
```

```
> influence.measures(fit)
Influence measures of
	 lm(formula = sr ~ ., data = LifeCycleSavings) :

                 dfb.1_ dfb.pp15 dfb.pp75  dfb.dpi  dfb.ddpi   dffit cov.r   cook.d    hat inf
Australia       0.01232 -0.01044 -0.02653  0.04534 -0.000159  0.0627 1.193 8.04e-04 0.0677    
Austria        -0.01005  0.00594  0.04084 -0.03672 -0.008182  0.0632 1.268 8.18e-04 0.1204    
Belgium        -0.06416  0.05150  0.12070 -0.03472 -0.007265  0.1878 1.176 7.15e-03 0.0875    
Bolivia         0.00578 -0.01270 -0.02253  0.03185  0.040642 -0.0597 1.224 7.28e-04 0.0895    
Brazil          0.08973 -0.06163 -0.17907  0.11997  0.068457  0.2646 1.082 1.40e-02 0.0696    
Canada          0.00541 -0.00675  0.01021 -0.03531 -0.002649 -0.0390 1.328 3.11e-04 0.1584    
Chile          -0.19941  0.13265  0.21979 -0.01998  0.120007 -0.4554 0.655 3.78e-02 0.0373   *
China           0.02112 -0.00573 -0.08311  0.05180  0.110627  0.2008 1.150 8.16e-03 0.0780    
Colombia        0.03910 -0.05226 -0.02464  0.00168  0.009084 -0.0960 1.167 1.88e-03 0.0573    
Costa Rica     -0.23367  0.28428  0.14243  0.05638 -0.032824  0.4049 0.968 3.21e-02 0.0755    
Denmark        -0.04051  0.02093  0.04653  0.15220  0.048854  0.3845 0.934 2.88e-02 0.0627    
Ecuador         0.07176 -0.09524 -0.06067  0.01950  0.047786 -0.1695 1.139 5.82e-03 0.0637    
Finland        -0.11350  0.11133  0.11695 -0.04364 -0.017132 -0.1464 1.203 4.36e-03 0.0920    
France         -0.16600  0.14705  0.21900 -0.02942  0.023952  0.2765 1.226 1.55e-02 0.1362    
Germany        -0.00802  0.00822  0.00835 -0.00697 -0.000293 -0.0152 1.226 4.74e-05 0.0874    
Greece         -0.14820  0.16394  0.02861  0.15713 -0.059599 -0.2811 1.140 1.59e-02 0.0966    
Guatamala       0.01552 -0.05485  0.00614  0.00585  0.097217 -0.2305 1.085 1.07e-02 0.0605    
Honduras       -0.00226  0.00984 -0.01020  0.00812 -0.001887  0.0482 1.186 4.74e-04 0.0601    
Iceland         0.24789 -0.27355 -0.23265 -0.12555  0.184698 -0.4768 0.866 4.35e-02 0.0705    
India           0.02105 -0.01577 -0.01439 -0.01374 -0.018958  0.0381 1.202 2.97e-04 0.0715    
Ireland        -0.31001  0.29624  0.48156 -0.25733 -0.093317  0.5216 1.268 5.44e-02 0.2122    
Italy           0.06619 -0.07097  0.00307 -0.06999 -0.028648  0.1388 1.162 3.92e-03 0.0665    
Japan           0.63987 -0.65614 -0.67390  0.14610  0.388603  0.8597 1.085 1.43e-01 0.2233    
Korea          -0.16897  0.13509  0.21895  0.00511 -0.169492 -0.4303 0.870 3.56e-02 0.0608    
Luxembourg     -0.06827  0.06888  0.04380 -0.02797  0.049134 -0.1401 1.196 3.99e-03 0.0863    
Malta           0.03652 -0.04876  0.00791 -0.08659  0.153014  0.2386 1.128 1.15e-02 0.0794    
Norway          0.00222 -0.00035 -0.00611 -0.01594 -0.001462 -0.0522 1.168 5.56e-04 0.0479    
Netherlands     0.01395 -0.01674 -0.01186  0.00433  0.022591  0.0366 1.229 2.74e-04 0.0906    
New Zealand    -0.06002  0.06510  0.09412 -0.02638 -0.064740  0.1469 1.134 4.38e-03 0.0542    
Nicaragua      -0.01209  0.01790  0.00972 -0.00474 -0.010467  0.0397 1.174 3.23e-04 0.0504    
Panama          0.02828 -0.05334  0.01446 -0.03467 -0.007889 -0.1775 1.067 6.33e-03 0.0390    
Paraguay       -0.23227  0.16416  0.15826  0.14361  0.270478 -0.4655 0.873 4.16e-02 0.0694    
Peru           -0.07182  0.14669  0.09148 -0.08585 -0.287184  0.4811 0.831 4.40e-02 0.0650    
Philippines    -0.15707  0.22681  0.15743 -0.11140 -0.170674  0.4884 0.818 4.52e-02 0.0643    
Portugal       -0.02140  0.02551 -0.00380  0.03991 -0.028011 -0.0690 1.233 9.73e-04 0.0971    
South Africa    0.02218 -0.02030 -0.00672 -0.02049 -0.016326  0.0343 1.195 2.41e-04 0.0651    
South Rhodesia  0.14390 -0.13472 -0.09245 -0.06956 -0.057920  0.1607 1.313 5.27e-03 0.1608    
Spain          -0.03035  0.03131  0.00394  0.03512  0.005340 -0.0526 1.208 5.66e-04 0.0773    
Sweden          0.10098 -0.08162 -0.06166 -0.25528 -0.013316 -0.4526 1.086 4.06e-02 0.1240    
Switzerland     0.04323 -0.04649 -0.04364  0.09093 -0.018828  0.1903 1.147 7.33e-03 0.0736    
Turkey         -0.01092 -0.01198  0.02645  0.00161  0.025138 -0.1445 1.100 4.22e-03 0.0396    
Tunisia         0.07377 -0.10500 -0.07727  0.04439  0.103058 -0.2177 1.131 9.56e-03 0.0746    
United Kingdom  0.04671 -0.03584 -0.17129  0.12554  0.100314 -0.2722 1.189 1.50e-02 0.1165    
United States   0.06910 -0.07289  0.03745 -0.23312 -0.032729 -0.2510 1.655 1.28e-02 0.3337   *
Venezuela      -0.05083  0.10080 -0.03366  0.11366 -0.124486  0.3071 1.095 1.89e-02 0.0863    
Zambia          0.16361 -0.07917 -0.33899  0.09406  0.228232  0.7482 0.512 9.66e-02 0.0643   *
Jamaica         0.10958 -0.10022 -0.05722 -0.00703 -0.295461 -0.3456 1.200 2.40e-02 0.1408    
Uruguay        -0.13403  0.12880  0.02953  0.13132  0.099591 -0.2051 1.187 8.53e-03 0.0979    
Libya           0.55074 -0.48324 -0.37974 -0.01937 -1.024477 -1.1601 2.091 2.68e-01 0.5315   *
Malaysia        0.03684 -0.06113  0.03235 -0.04956 -0.072294 -0.2126 1.113 9.11e-03 0.0652    
```

Here, the hat column contains the hat values i.e., diagonal entries of hat matrix $X(XX')^{-1}X'$.

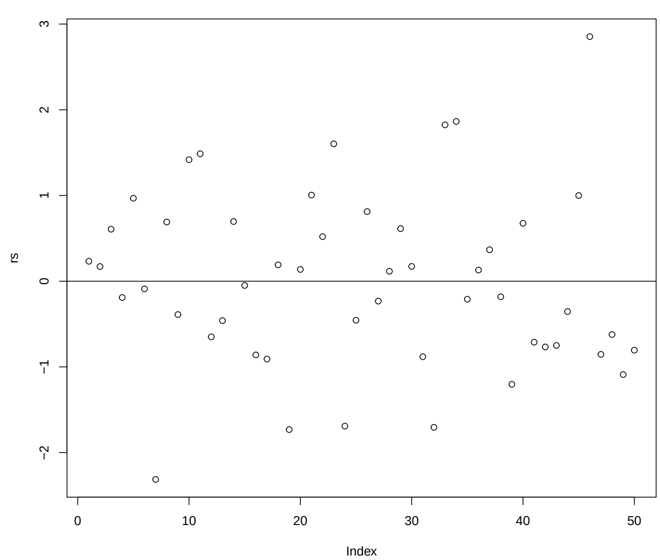
Now based on all these columns we get here, we have to conclude which are the influential cases. There are various cut off values of these parameters by which we conclude. For more information on cut off values and their justifications follow **Besley Kuh Welsch**.

Here we get a column named *inf* from which we directly get those points which are influential. Eg.- here Chile, United States, Zambia, Libya are four influential points. So they need another law, but they have to be in our data, we also have to check if there is any typo or not in recording these influential points.

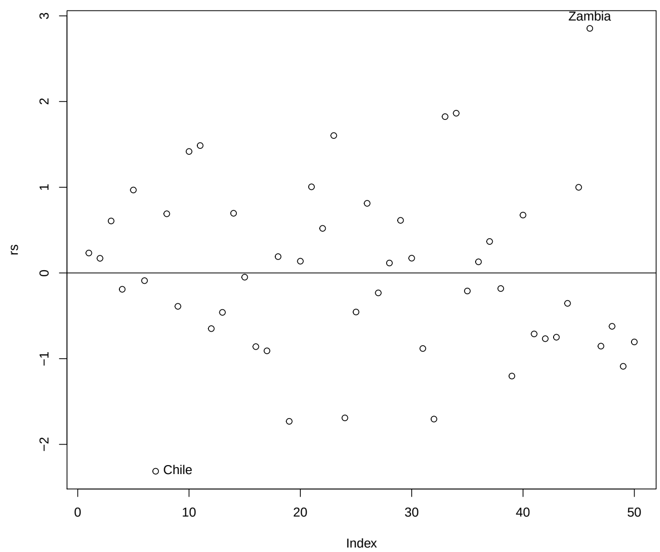
Now we may want to check the parameters to detect influential points individually. So, we can find out the r-student values by using the function rstudent().

```
rs=rstudent(fit)
plot(rs);
abline(h=0)
identify(rs, lab= rownames(LifeCycleSavings))
```

Now we plot these r-student values with the line rs=0. The plot is shown below:



From r-student values we can identify the outliers only, it doesn't tell us if a point is influential or not. Now from the plot we can identify that there are two outlier points plotted on the graph, and to know which points are these we use the function *identify()*. It takes the plotted values and labels associated with those plotted points as its input. By using this function we can locate the outlier points in our plot by selecting the points using the cursor(which changes into locator) and we get the names associated, thus we can identify these points. The plot after identifying these points is shown below:

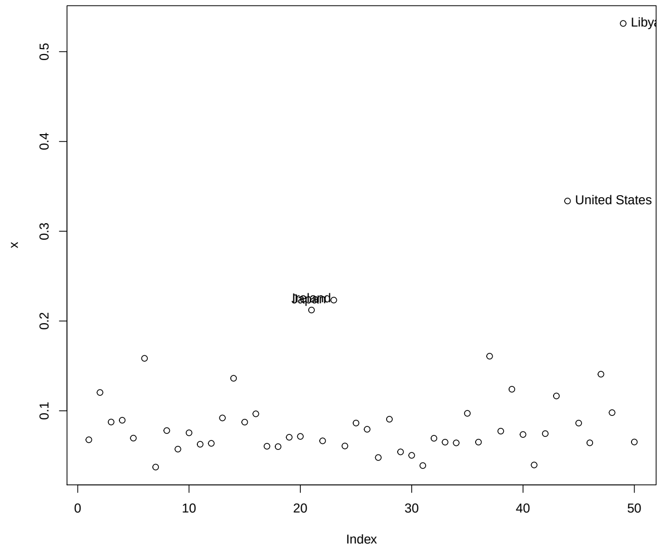


So, by identifying the outliers we get two points which are Zambia and Chile which we got as influential points using the function *influence.measure()*. So, out of four influential points two are outliers and the remaining points are not outliers.

```
checkplot= function(x,lab){
  plot(x)
  identify(x, lab=lab)
}

checkplot(hatvalues(fit),rownames(LifeCycleSavings))
```

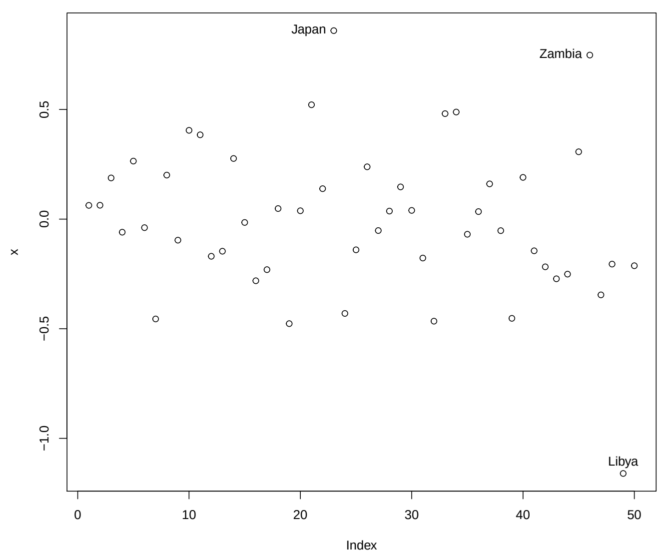
Now let's take a look at the hat-values. To do the plots easily we write a function *checkplot()* which takes the value to be plotted and the names associated with those points as inputs and gives the plot and an option to identify points on the plot as output. So using this function we plot hat-values and identify those points which don't behave like the other points behave. The plot is given below:



From the plot we report that Libya and USA are some special points i.e., they don't behave like the bulk of the data. And we also notice that Japan and Ireland also differs a little bit from the general behaviour of other data-points so we may consider these two points separately.

```
checkplot(dffits(fit),rownames(LifeCycleSavings))
```

Now let's look at dffits. By the same way we plot the values of dffits. The plot is given below:



From this plot we get Libya, Zambia and Japan as influential points, where Libya and Zambia are already reported.

So, this is how we carry out influence analysis. But above all we should remember that by doing this analysis we can identify those certain points which are potentially dangerous. Now at last we should check if there is any typo in recording the values of these points, we should check these values with the domain knowledge etc.