INFLUENCE DIAGNOSTICS
R LAB

RAHUL DEBNATH
Under mentorship of Prof. Arnab Chakraborty

21 OCTOBER 2022

Indian Statistical Institute, Kolkata.

"THE FUTURE BELONGS TO THOSE WHO BELIEVE IN THE BEAUTY OF THEIR DREAMS."

Eleanor Roosevelt

Hello there, readers ! Yes, you. I need your eyes and attention here in 3 ...... 2 .... 1 ......

Today, what we are interested to do is learning how to do influence diagnostics using R.

For this we are going to use the dataset present in R which is known as Life Cycle Savings Data which is a dataset from economics. So, lets quickly have a look at the dataset in R using the command :

LifeCycleSavings

We get the output as :

```
                sr pop15 pop75     dpi  ddpi
Australia      11.43 29.35  2.87 2329.68  2.87
Austria        12.07 23.32  4.41 1507.99  3.93
Belgium        13.17 23.80  4.43 2108.47  3.82
Bolivia         5.75 41.89  1.67  189.13  0.22
Brazil         12.88 42.19  0.83  728.47  4.56
Canada          8.79 31.72  2.85 2982.88  2.43
Chile           0.60 39.74  1.34  662.86  2.67
China          11.90 44.75  0.67  289.52  6.51
Colombia        4.98 46.64  1.06  276.65  3.08
Costa Rica     10.78 47.64  1.14  471.24  2.80
Denmark        16.85 24.42  3.93 2496.53  3.99
Ecuador         3.59 46.31  1.19  287.77  2.19
Finland        11.24 27.84  2.37 1681.25  4.32
France         12.64 25.06  4.70 2213.82  4.52
Germany        12.55 23.31  3.35 2457.12  3.44
Greece         10.67 25.62  3.10  870.85  6.28
Guatamala       3.01 46.05  0.87  289.71  1.48
Honduras        7.70 47.32  0.58  232.44  3.19
Iceland         1.27 34.03  3.08 1900.10  1.12
India           9.00 41.31  0.96   88.94  1.54
Ireland        11.34 31.16  4.19 1139.95  2.99
Italy          14.28 24.52  3.48 1390.00  3.54
Japan          21.10 27.01  1.91 1257.28  8.21
Korea           3.98 41.74  0.91  207.68  5.81
Luxembourg     10.35 21.80  3.73 2449.39  1.57
Malta          15.48 32.54  2.47  601.05  8.12
Norway         10.25 25.95  3.67 2231.03  3.62
Netherlands    14.65 24.71  3.25 1740.70  7.66
New Zealand    10.67 32.61  3.17 1487.52  1.76
Nicaragua       7.30 45.04  1.21  325.54  2.48
Panama          4.44 43.56  1.20  568.56  3.61
Paraguay        2.02 41.18  1.05  220.56  1.03
Peru           12.70 44.19  1.28  400.06  0.67
```

```
Philippines    12.78 46.26  1.12  152.01  2.00
Portugal       12.49 28.96  2.85  579.51  7.48
South Africa   11.14 31.94  2.28  651.11  2.19
South Rhodesia 13.30 31.92  1.52  250.96  2.00
Spain          11.77 27.74  2.87  768.79  4.35
Sweden          6.86 21.44  4.54 3299.49  3.01
Switzerland    14.13 23.49  3.73 2630.96  2.70
Turkey          5.13 43.42  1.08  389.66  2.96
Tunisia         2.81 46.12  1.21  249.87  1.13
United Kingdom  7.81 23.27  4.46 1813.93  2.01
United States   7.56 29.81  3.43 4001.89  2.45
Venezuela       9.22 46.40  0.90  813.39  0.53
Zambia         18.56 45.25  0.56  138.33  5.14
Jamaica         7.72 41.12  1.73  380.47 10.23
Uruguay         9.24 28.13  2.72  766.54  1.88
Libya           8.89 43.69  2.07  123.58 16.71
Malaysia        4.71 47.20  0.66  242.69  5.08
```

Now, we use the following command to know about the dataset :

?LifeCycleSavings

What we get to know from the output is the following :

```
LifeCycleSavings {datasets} R Documentation
Intercountry Life-Cycle Savings Data
Description
Data on the savings ratio 1960{1970.

Usage
LifeCycleSavings
Format
A data frame with 50 observations on 5 variables.

[,1] sr numeric aggregate personal savings
[,2] pop15 numeric % of population under 15
[,3] pop75 numeric % of population over 75
[,4] dpi numeric real per-capita disposable income
[,5] ddpi numeric % growth rate of dpi
Details
Under the life-cycle savings hypothesis as developed by Franco Modigliani,
 the savings ratio (aggregate personal saving divided by disposable income)
 is explained by per-capita disposable income, the percentage rate of change
 in per-capita disposable income, and two demographic variables:
 the percentage of population less than 15 years old and the percentage of the
 population over 75 years old. The data are averaged over the decade 1960{1970
 to remove the business cycle or other short-term fluctuations.

Source
The data were obtained from Belsley, Kuh and Welsch (1980).
They in turn obtained the data from Sterling (1977).
```

```
References
Sterling, Arnie (1977) Unpublished BS Thesis. Massachusetts Institute of Technology.

Belsley, D. A., Kuh. E. and Welsch, R. E. (1980)
 Regression Diagnostics. New York: Wiley.

Examples
require(stats); require(graphics)
pairs(LifeCycleSavings, panel = panel.smooth,
      main = "LifeCycleSavings data")
fm1 <- lm(sr ~ pop15 + pop75 + dpi + ddpi, data = LifeCycleSavings)
summary(fm1)
```

So, we notice that there are a number of countries which constitute the basic unit of our dataset and for each countries we have some measures of different variables. Out of all the variables 'sr' is the response and the remaining varibles are the predictors. Now to know the names of all the countries listed, we use the following R command:

rownames(LifeCycleSavings)

So, as an output we get all the names of the countries as :

```
 [1] "Australia"      "Austria"        "Belgium"        "Bolivia"
 [5] "Brazil"         "Canada"         "Chile"          "China"
 [9] "Colombia"       "Costa Rica"     "Denmark"        "Ecuador"
[13] "Finland"        "France"         "Germany"        "Greece"
[17] "Guatamala"      "Honduras"       "Iceland"        "India"
[21] "Ireland"        "Italy"          "Japan"          "Korea"
[25] "Luxembourg"     "Malta"          "Norway"         "Netherlands"
[29] "New Zealand"    "Nicaragua"      "Panama"         "Paraguay"
[33] "Peru"           "Philippines"    "Portugal"       "South Africa"
[37] "South Rhodesia" "Spain"          "Sweden"         "Switzerland"
[41] "Turkey"         "Tunisia"        "United Kingdom" "United States"
[45] "Venezuela"      "Zambia"         "Jamaica"        "Uruguay"
[49] "Libya"          "Malaysia"
```

We can see the names of all the 50 countries now.This data is averaged over a 10 year period time and based on this we want to ckeck the model known as life-cycle-savings hypothesis . For doing this 'sr' is used as the response and the remaining four as the predictors.
Now, we use the following R command to kn0w the names:

names(LifeCycleSavings)

As an output we have:

```
[1] "sr"    "pop15" "pop75" "dpi"   "ddpi"
```

Now we fit the model using the command:

fit = lm(sr .,data = LifeCycleSavings)

The dot means the rest four variables as a sum. Now, to view the summary we use the command:

summary(fit)

The output is :

```
Call:
lm(formula = sr ~ ., data = LifeCycleSavings)

Residuals:
    Min      1Q  Median      3Q     Max
-8.2422 -2.6857 -0.2488  2.4280  9.7509

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 28.5660865  7.3545161   3.884 0.000334 ***
pop15       -0.4611931  0.1446422  -3.189 0.002603 **
pop75       -1.6914977  1.0835989  -1.561 0.125530
dpi         -0.0003369  0.0009311  -0.362 0.719173
ddpi         0.4096949  0.1961971   2.088 0.042471 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.803 on 45 degrees of freedom
Multiple R-squared:  0.3385,    Adjusted R-squared:  0.2797
F-statistic: 5.756 on 4 and 45 DF,  p-value: 0.0007904
```

Here, we notice that the values of multiple R-squared and adjusted R-squared are quite mischiviously low but our book remarks that these values are completely fine as we cannot provide any absolute threshold for their values, since they are very much domain specific. For this case it is claimed to be good enough to work with these values.Now, to know how many observations we have, we use the command:

nrows(LifeCycleSavings)

Output:

```
[1] 50
```

Now, we need to know out of these 50 countries if something is weird.
But the question is how can we know that? Can you think a bit ?

Well, there is a nice function in R which will solve everything for us popularly known as influence-measures. In R, we just run the command :

influence.measures(fit)

We get the output :

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

                 dfb.1_ dfb.pp15 dfb.pp75  dfb.dpi  dfb.ddpi   dffit cov.r
```

Australia	0.01232	-0.01044	-0.02653	0.04534	-0.000159	0.0627	1.193
Austria	-0.01005	0.00594	0.04084	-0.03672	-0.008182	0.0632	1.268
Belgium	-0.06416	0.05150	0.12070	-0.03472	-0.007265	0.1878	1.176
Bolivia	0.00578	-0.01270	-0.02253	0.03185	0.040642	-0.0597	1.224
Brazil	0.08973	-0.06163	-0.17907	0.11997	0.068457	0.2646	1.082
Canada	0.00541	-0.00675	0.01021	-0.03531	-0.002649	-0.0390	1.328
Chile	-0.19941	0.13265	0.21979	-0.01998	0.120007	-0.4554	0.655
China	0.02112	-0.00573	-0.08311	0.05180	0.110627	0.2008	1.150
Colombia	0.03910	-0.05226	-0.02464	0.00168	0.009084	-0.0960	1.167
Costa Rica	-0.23367	0.28428	0.14243	0.05638	-0.032824	0.4049	0.968
Denmark	-0.04051	0.02093	0.04653	0.15220	0.048854	0.3845	0.934
Ecuador	0.07176	-0.09524	-0.06067	0.01950	0.047786	-0.1695	1.139
Finland	-0.11350	0.11133	0.11695	-0.04364	-0.017132	-0.1464	1.203
France	-0.16600	0.14705	0.21900	-0.02942	0.023952	0.2765	1.226
Germany	-0.00802	0.00822	0.00835	-0.00697	-0.000293	-0.0152	1.226
Greece	-0.14820	0.16394	0.02861	0.15713	-0.059599	-0.2811	1.140
Guatamala	0.01552	-0.05485	0.00614	0.00585	0.097217	-0.2305	1.085
Honduras	-0.00226	0.00984	-0.01020	0.00812	-0.001887	0.0482	1.186
Iceland	0.24789	-0.27355	-0.23265	-0.12555	0.184698	-0.4768	0.866
India	0.02105	-0.01577	-0.01439	-0.01374	-0.018958	0.0381	1.202
Ireland	-0.31001	0.29624	0.48156	-0.25733	-0.093317	0.5216	1.268
Italy	0.06619	-0.07097	0.00307	-0.06999	-0.028648	0.1388	1.162
Japan	0.63987	-0.65614	-0.67390	0.14610	0.388603	0.8597	1.085
Korea	-0.16897	0.13509	0.21895	0.00511	-0.169492	-0.4303	0.870
Luxembourg	-0.06827	0.06888	0.04380	-0.02797	0.049134	-0.1401	1.196
Malta	0.03652	-0.04876	0.00791	-0.08659	0.153014	0.2386	1.128
Norway	0.00222	-0.00035	-0.00611	-0.01594	-0.001462	-0.0522	1.168
Netherlands	0.01395	-0.01674	-0.01186	0.00433	0.022591	0.0366	1.229
New Zealand	-0.06002	0.06510	0.09412	-0.02638	-0.064740	0.1469	1.134
Nicaragua	-0.01209	0.01790	0.00972	-0.00474	-0.010467	0.0397	1.174
Panama	0.02828	-0.05334	0.01446	-0.03467	-0.007889	-0.1775	1.067
Paraguay	-0.23227	0.16416	0.15826	0.14361	0.270478	-0.4655	0.873
Peru	-0.07182	0.14669	0.09148	-0.08585	-0.287184	0.4811	0.831
Philippines	-0.15707	0.22681	0.15743	-0.11140	-0.170674	0.4884	0.818
Portugal	-0.02140	0.02551	-0.00380	0.03991	-0.028011	-0.0690	1.233
South Africa	0.02218	-0.02030	-0.00672	-0.02049	-0.016326	0.0343	1.195
South Rhodesia	0.14390	-0.13472	-0.09245	-0.06956	-0.057920	0.1607	1.313
Spain	-0.03035	0.03131	0.00394	0.03512	0.005340	-0.0526	1.208
Sweden	0.10098	-0.08162	-0.06166	-0.25528	-0.013316	-0.4526	1.086
Switzerland	0.04323	-0.04649	-0.04364	0.09093	-0.018828	0.1903	1.147
Turkey	-0.01092	-0.01198	0.02645	0.00161	0.025138	-0.1445	1.100
Tunisia	0.07377	-0.10500	-0.07727	0.04439	0.103058	-0.2177	1.131
United Kingdom	0.04671	-0.03584	-0.17129	0.12554	0.100314	-0.2722	1.189
United States	0.06910	-0.07289	0.03745	-0.23312	-0.032729	-0.2510	1.655
Venezuela	-0.05083	0.10080	-0.03366	0.11366	-0.124486	0.3071	1.095
Zambia	0.16361	-0.07917	-0.33899	0.09406	0.228232	0.7482	0.512
Jamaica	0.10958	-0.10022	-0.05722	-0.00703	-0.295461	-0.3456	1.200
Uruguay	-0.13403	0.12880	0.02953	0.13132	0.099591	-0.2051	1.187
Libya	0.55074	-0.48324	-0.37974	-0.01937	-1.024477	-1.1601	2.091
Malaysia	0.03684	-0.06113	0.03235	-0.04956	-0.072294	-0.2126	1.113

	cook.d	hat	inf
Australia	8.04e-04	0.0677	
Austria	8.18e-04	0.1204	
Belgium	7.15e-03	0.0875	
Bolivia	7.28e-04	0.0895	
Brazil	1.40e-02	0.0696	
Canada	3.11e-04	0.1584	
Chile	3.78e-02	0.0373	*
China	8.16e-03	0.0780	
Colombia	1.88e-03	0.0573	
Costa Rica	3.21e-02	0.0755	
Denmark	2.88e-02	0.0627	
Ecuador	5.82e-03	0.0637	
Finland	4.36e-03	0.0920	
France	1.55e-02	0.1362	
Germany	4.74e-05	0.0874	
Greece	1.59e-02	0.0966	
Guatamala	1.07e-02	0.0605	
Honduras	4.74e-04	0.0601	
Iceland	4.35e-02	0.0705	
India	2.97e-04	0.0715	
Ireland	5.44e-02	0.2122	
Italy	3.92e-03	0.0665	
Japan	1.43e-01	0.2233	
Korea	3.56e-02	0.0608	
Luxembourg	3.99e-03	0.0863	
Malta	1.15e-02	0.0794	
Norway	5.56e-04	0.0479	
Netherlands	2.74e-04	0.0906	
New Zealand	4.38e-03	0.0542	
Nicaragua	3.23e-04	0.0504	
Panama	6.33e-03	0.0390	
Paraguay	4.16e-02	0.0694	
Peru	4.40e-02	0.0650	
Philippines	4.52e-02	0.0643	
Portugal	9.73e-04	0.0971	
South Africa	2.41e-04	0.0651	
South Rhodesia	5.27e-03	0.1608	
Spain	5.66e-04	0.0773	
Sweden	4.06e-02	0.1240	
Switzerland	7.33e-03	0.0736	
Turkey	4.22e-03	0.0396	
Tunisia	9.56e-03	0.0746	
United Kingdom	1.50e-02	0.1165	
United States	1.28e-02	0.3337	*
Venezuela	1.89e-02	0.0863	
Zambia	9.66e-02	0.0643	*
Jamaica	2.40e-02	0.1408	
Uruguay	8.53e-03	0.0979	
Libya	2.68e-01	0.5315	*

```
Malaysia       9.11e-03 0.0652
```

For each value of i, i.e.,in our case for each country, it computes the values of the DFBETA(S) i.e., for the intercept ($dfb.1_)$ and four regressors($dfb.pp15, dfb.pp75, dfb.dpianddfb.ddpi$), DFFIT i.e.,($dffit$) and the cov ratios i.e.,($cov.r$). There are some other values which doesnot fit in the table, so they are written below in next line, which contains values of cook's distance (cook.d).

That's all for this session,we shall continue with knowing what is Cook's distance in our next session.

Thank You for reading with patience.

NOTE : The R code is also submitted as .R file.

The R code used :

```
LifeCycleSavings
?LifeCycleSavings
rownames(LifeCycleSavings)
names(LifeCycleSavings)
fit=lm(sr~.,data=LifeCycleSavings)
summary(fit)
nrows(LifeCycleSavings)
influence.measures(fit)
```