Regression Analysis

Abith Raj

2023-08-29

```
# Regression Analysis
swiss  # This is the inbuilt dataset
```

```
##              Fertility Agriculture Examination Education Catholic
## Courtelary        80.2        17.0          15        12     9.96
## Delemont          83.1        45.1           6         9    84.84
## Franches-Mnt      92.5        39.7           5         5    93.40
## Moutier           85.8        36.5          12         7    33.77
## Neuveville        76.9        43.5          17        15     5.16
## Porrentruy        76.1        35.3           9         7    90.57
## Broye             83.8        70.2          16         7    92.85
## Glane             92.4        67.8          14         8    97.16
## Gruyere           82.4        53.3          12         7    97.67
## Sarine            82.9        45.2          16        13    91.38
## Veveyse           87.1        64.5          14         6    98.61
## Aigle             64.1        62.0          21        12     8.52
## Aubonne           66.9        67.5          14         7     2.27
## Avenches          68.9        60.7          19        12     4.43
## Cossonay          61.7        69.3          22         5     2.82
## Echallens         68.3        72.6          18         2    24.20
## Grandson          71.7        34.0          17         8     3.30
## Lausanne          55.7        19.4          26        28    12.11
## La Vallee         54.3        15.2          31        20     2.15
## Lavaux            65.1        73.0          19         9     2.84
## Morges            65.5        59.8          22        10     5.23
## Moudon            65.0        55.1          14         3     4.52
## Nyone             56.6        50.9          22        12    15.14
## Orbe              57.4        54.1          20         6     4.20
## Oron              72.5        71.2          12         1     2.40
## Payerne           74.2        58.1          14         8     5.23
## Paysd'enhaut      72.0        63.5           6         3     2.56
## Rolle             60.5        60.8          16        10     7.72
## Vevey             58.3        26.8          25        19    18.46
## Yverdon           65.4        49.5          15         8     6.10
## Conthey           75.5        85.9           3         2    99.71
## Entremont         69.3        84.9           7         6    99.68
## Herens            77.3        89.7           5         2   100.00
## Martigwy          70.5        78.2          12         6    98.96
## Monthey           79.4        64.9           7         3    98.22
## St Maurice        65.0        75.9           9         9    99.06
## Sierre            92.2        84.6           3         3    99.46
## Sion              79.3        63.1          13        13    96.83
## Boudry            70.4        38.4          26        12     5.62
## La Chauxdfnd      65.7         7.7          29        11    13.79
## Le Locle          72.7        16.7          22        13    11.22
## Neuchatel         64.4        17.6          35        32    16.92
## Val de Ruz        77.6        37.6          15         7     4.97
## ValdeTravers      67.6        18.7          25         7     8.65
## V. De Geneve      35.0         1.2          37        53    42.34
## Rive Droite       44.7        46.6          16        29    50.43
## Rive Gauche       42.8        27.7          22        29    58.33
##              Infant.Mortality
## Courtelary               22.2
## Delemont                 22.2
## Franches-Mnt             20.2
## Moutier                  20.3
## Neuveville               20.6
## Porrentruy               26.6
```

```
## Broye                    23.6
## Glane                    24.9
## Gruyere                  21.0
## Sarine                   24.4
## Veveyse                  24.5
## Aigle                    16.5
## Aubonne                  19.1
## Avenches                 22.7
## Cossonay                 18.7
## Echallens                21.2
## Grandson                 20.0
## Lausanne                 20.2
## La Vallee                10.8
## Lavaux                   20.0
## Morges                   18.0
## Moudon                   22.4
## Nyone                    16.7
## Orbe                     15.3
## Oron                     21.0
## Payerne                  23.8
## Paysd'enhaut             18.0
## Rolle                    16.3
## Vevey                    20.9
## Yverdon                  22.5
## Conthey                  15.1
## Entremont                19.8
## Herens                   18.3
## Martigwy                 19.4
## Monthey                  20.2
## St Maurice               17.8
## Sierre                   16.3
## Sion                     18.1
## Boudry                   20.3
## La Chauxdfnd             20.5
## Le Locle                 18.9
## Neuchatel                23.0
## Val de Ruz               20.0
## ValdeTravers             19.5
## V. De Geneve             18.0
## Rive Droite              18.2
## Rive Gauche              19.3
```

```
dim(swiss) # Checking dimension
```

```
## [1] 47  6
```

47 rows, 6 columns

```
head(swiss) # Displays only 5 rows in the dataset
```

```
##              Fertility Agriculture Examination Education Catholic
## Courtelary        80.2        17.0          15        12     9.96
## Delemont          83.1        45.1           6         9    84.84
## Franches-Mnt      92.5        39.7           5         5    93.40
## Moutier           85.8        36.5          12         7    33.77
## Neuveville        76.9        43.5          17        15     5.16
## Porrentruy        76.1        35.3           9         7    90.57
##              Infant.Mortality
## Courtelary               22.2
## Delemont                 22.2
## Franches-Mnt             20.2
## Moutier                  20.3
## Neuveville               20.6
## Porrentruy               26.6
```

Dataset that contains socio-economic indicators for various provinces in Switzerland. It's often used for educational purposes and to demonstrate different statistical techniques.

```
crls=cor(swiss)  # Checking correrlation
crls
```

```
##                  Fertility Agriculture Examination   Education   Catholic
## Fertility         1.0000000  0.35307918  -0.6458827 -0.66378886  0.4636847
## Agriculture       0.3530792  1.00000000  -0.6865422 -0.63952252  0.4010951
## Examination      -0.6458827 -0.68654221   1.0000000  0.69841530 -0.5727418
## Education        -0.6637889 -0.63952252   0.6984153  1.00000000 -0.1538589
## Catholic          0.4636847  0.40109505  -0.5727418 -0.15385892  1.0000000
## Infant.Mortality  0.4165560 -0.06085861  -0.1140216 -0.09932185  0.1754959
##                  Infant.Mortality
## Fertility              0.41655603
## Agriculture           -0.06085861
## Examination           -0.11402160
## Education             -0.09932185
## Catholic               0.17549591
## Infant.Mortality       1.00000000
```

```
round(crls,2) # Making it has 2 decimal numbers
```

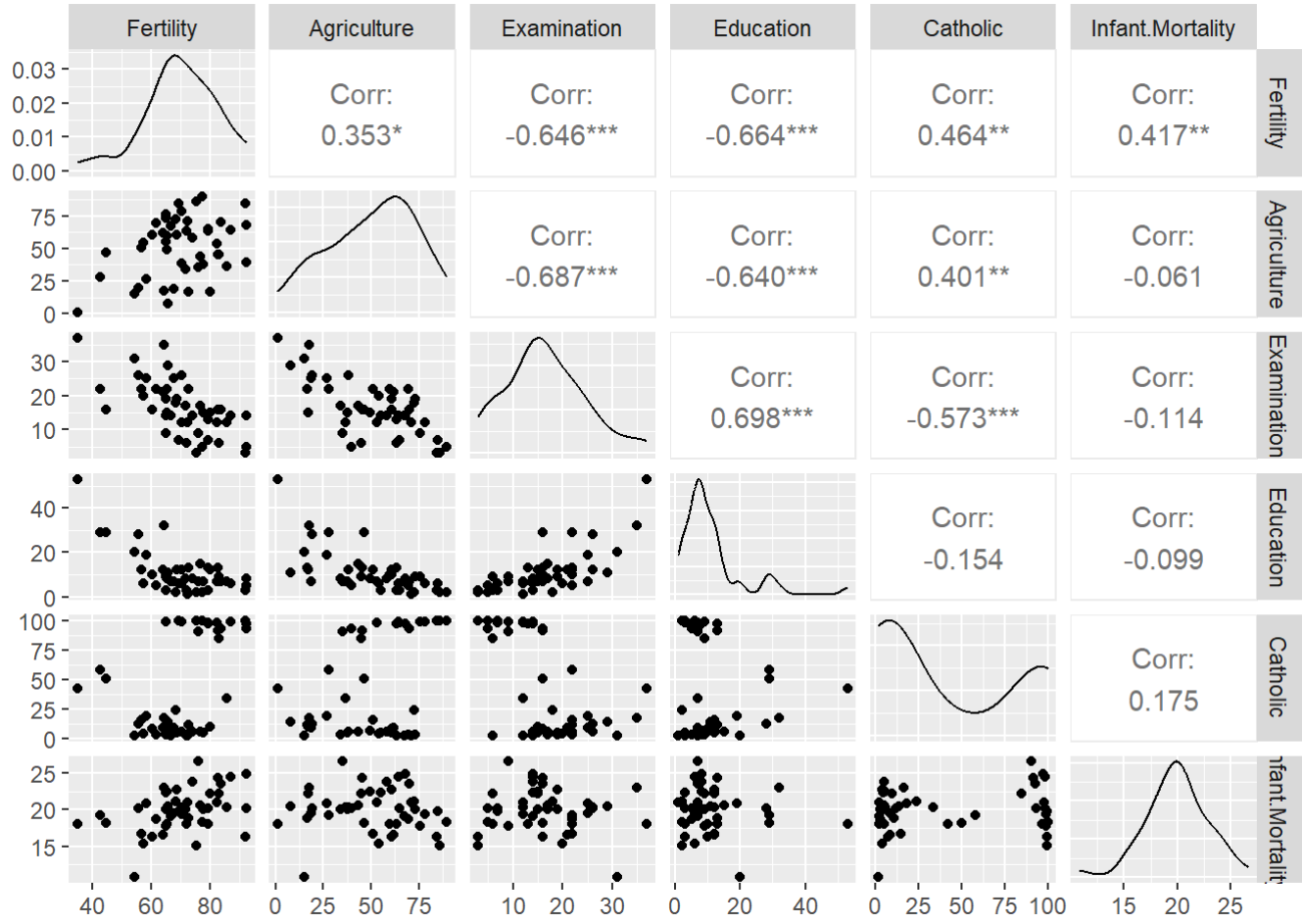
```
##                 Fertility Agriculture Examination Education Catholic
## Fertility            1.00        0.35       -0.65     -0.66     0.46
## Agriculture          0.35        1.00       -0.69     -0.64     0.40
## Examination         -0.65       -0.69        1.00      0.70    -0.57
## Education           -0.66       -0.64        0.70      1.00    -0.15
## Catholic             0.46        0.40       -0.57     -0.15     1.00
## Infant.Mortality     0.42       -0.06       -0.11     -0.10     0.18
##                 Infant.Mortality
## Fertility                   0.42
## Agriculture                -0.06
## Examination                -0.11
## Education                  -0.10
## Catholic                    0.18
## Infant.Mortality            1.00
```

```
library(GGally)
```

```
## Loading required package: ggplot2
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
ggpairs(swiss)
```



In this education and examination are highly correlated that is 0.7. Education and agriculture are highly negatively correlated.

```
attach(swiss)
mlr=lm(Fertility~.,data=swiss) # Performing multilinear regression
mlr
```

```
##
## Call:
## lm(formula = Fertility ~ ., data = swiss)
##
## Coefficients:
##      (Intercept)       Agriculture       Examination         Education
##          66.9152           -0.1721           -0.2580           -0.8709
##         Catholic  Infant.Mortality
##           0.1041            1.0770
```

```
summary(mlr) # Summary of the MLr
```

```
##
## Call:
## lm(formula = Fertility ~ ., data = swiss)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.2743  -5.2617   0.5032   4.1198  15.3213
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      66.91518   10.70604   6.250 1.91e-07 ***
## Agriculture      -0.17211    0.07030  -2.448  0.01873 *
## Examination      -0.25801    0.25388  -1.016  0.31546
## Education        -0.87094    0.18303  -4.758 2.43e-05 ***
## Catholic          0.10412    0.03526   2.953  0.00519 **
## Infant.Mortality  1.07705    0.38172   2.822  0.00734 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.165 on 41 degrees of freedom
## Multiple R-squared:  0.7067, Adjusted R-squared:  0.671
## F-statistic: 19.76 on 5 and 41 DF,  p-value: 5.594e-10
```

```
anova(mlr) # ANOVA test
```

```
## Analysis of Variance Table
##
## Response: Fertility
##                  Df  Sum Sq Mean Sq F value    Pr(>F)
## Agriculture       1  894.84  894.84 17.4288 0.0001515 ***
## Examination       1 2210.38 2210.38 43.0516 6.885e-08 ***
## Education         1  891.81  891.81 17.3699 0.0001549 ***
## Catholic          1  667.13  667.13 12.9937 0.0008387 ***
## Infant.Mortality  1  408.75  408.75  7.9612 0.0073357 **
## Residuals        41 2105.04   51.34
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this dataset all the variables are highly significant.

Now all variaables are significant

```
#Assumptions
e=residuals(mlr)
e
```

```
##   Courtelary     Delemont Franches-Mnt      Moutier   Neuveville   Porrentruy
##    5.5847028    0.5900550    6.5817401    8.9796056   12.1975903  -14.4001100
##        Broye        Glane      Gruyere       Sarine      Veveyse        Aigle
##    4.1064868   10.7994372    1.0642170    3.4206889    3.4694330    5.0669814
##      Aubonne     Avenches     Cossonay    Echallens     Grandson     Lausanne
##    0.5032459    2.8753500   -3.6912074   -5.0866905    0.1058105    0.2011563
##    La Vallee       Lavaux       Morges       Moudon        Nyone         Orbe
##    3.5620395    1.6531050    3.3314260  -10.8032460   -4.9901286   -6.7341316
##         Oron      Payerne Paysd'enhaut        Rolle        Vevey      Yverdon
##   -1.0615158    1.6860070    0.5215233   -1.4727748   -5.4367352   -7.0265827
##      Conthey    Entremont       Herens     Martigwy      Monthey   St Maurice
##   -0.7594533   -7.6747768   -1.2661512   -5.8321170   -3.9086873   -8.1763181
##       Sierre         Sion       Boudry La Chauxdfnd     Le Locle    Neuchatel
##   15.3213097    8.3454790    4.8042844   -6.3425618    4.1331399   10.8806560
##   Val de Ruz ValdeTravers V. De Geneve  Rive Droite  Rive Gauche
##    5.0645917   -5.4529003    0.2023737   -9.6620893  -15.2742579
```

```
sum(e)
```

```
## [1] -2.442491e-15
```

```
round(sum(e),5)
```

```
## [1] 0
```

The sum of error terms is 0.

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(car)
```

```
## Loading required package: carData
```

```
vif_values <- vif(mlr)
print(vif_values)
```

```
##      Agriculture      Examination        Education         Catholic
##         2.284129         3.675420         2.774943         1.937160
## Infant.Mortality
##         1.107542
```

```
#durbin watson test for auto correlation
dwtest(mlr)
```

```
##
##  Durbin-Watson test
##
## data:  mlr
## DW = 1.4535, p-value = 0.01131
## alternative hypothesis: true autocorrelation is greater than 0
```

```
# bp test for checking homoscedasticity  (variance of the diagonals is equal or not  )
bptest(mlr)
```

```
##
##  studentized Breusch-Pagan test
##
## data:  mlr
## BP = 5.8511, df = 5, p-value = 0.321
```

```
#Stocastic Test
library(Matrix)
a=data.frame(Agriculture,Examination,Education,Catholic,Infant.Mortality)
rankMatrix(as.matrix(a))
```

```
## [1] 5
## attr(,"method")
## [1] "tolNorm2"
## attr(,"useGrad")
## [1] FALSE
## attr(,"tol")
## [1] 1.04361e-14
```

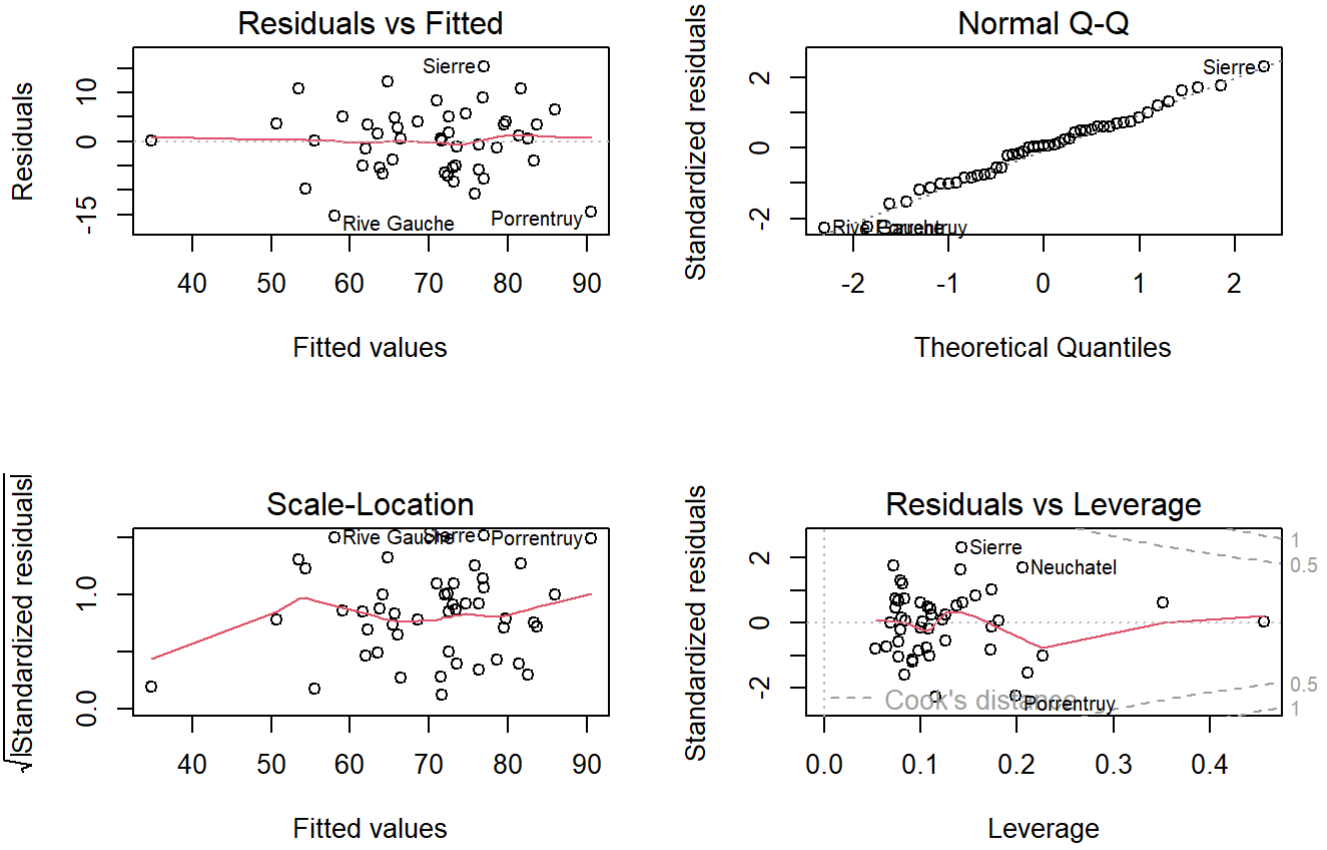
The number of variables is equal to the number of rank.

```
# Cheking for normality of residuals
shapiro.test(e)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  e
## W = 0.98892, p-value = 0.9318
```

Shapira test is used to check whether the data follow normal distribution or not. In this dataset it follows normal distribution as it its p-value is more 0.05. Accept the null hypothesis.

```
par(mfrow=c(2,2))
plot(mlr)
```




```
# Calculate 95% confidence intervals
confint_95 <- confint(mlr,level = 0.95)
# Print the confidence intervals
print(confint_95)
```

```
##                       2.5 %      97.5 %
## (Intercept)      45.29390014 88.53646321
## Agriculture      -0.31409562 -0.03013232
## Examination      -0.77072567  0.25470919
## Education        -1.24057382 -0.50130630
## Catholic          0.03291065  0.17532001
## Infant.Mortality  0.30614967  1.84794661
```

```r
# Calculate 99% confidence intervals
confint_99 <- confint(mlr, level = 0.99)
print(confint_99)
```

```
##                         0.5 %      99.5 %
## (Intercept)      37.996233118 95.83413024
## Agriculture      -0.362017614  0.01778967
## Examination      -0.943779289  0.42776281
## Education        -1.365333500 -0.37654663
## Catholic          0.008877479  0.19935318
## Infant.Mortality  0.045954157  2.10814212
```