Multi-colliniearity-Longley

SP

28/10/2019

```
library(car)
```

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

```
library(MASS)
```

```
ds<-read.csv("longley.csv")
ds
```

```
##        GNP GNP.deflator Unemployed Armed.Forces Population Year Employed
## 1  234.289         83.0      235.6        159.0    107.608 1947   60.323
## 2  259.426         88.5      232.5        145.6    108.632 1948   61.122
## 3  258.054         88.2      368.2        161.6    109.773 1949   60.171
## 4  284.599         89.5      335.1        165.0    110.929 1950   61.187
## 5  328.975         96.2      209.9        309.9    112.075 1951   63.221
## 6  346.999         98.1      193.2        359.4    113.270 1952   63.639
## 7  365.385         99.0      187.0        354.7    115.094 1953   64.989
## 8  363.112        100.0      357.8        335.0    116.219 1954   63.761
## 9  397.469        101.2      290.4        304.8    117.388 1955   66.019
## 10 419.180        104.6      282.2        285.7    118.734 1956   67.857
## 11 442.769        108.4      293.6        279.8    120.445 1957   68.169
## 12 444.546        110.8      468.1        263.7    121.950 1958   66.513
## 13 482.704        112.6      381.3        255.2    123.366 1959   68.655
## 14 502.601        114.2      393.1        251.4    125.368 1960   69.564
## 15 518.173        115.7      480.6        257.2    127.852 1961   69.331
## 16 554.894        116.9      400.7        282.7    130.081 1962   70.551
```

Model

```
lm=lm(Employed ~ ., data=ds)
summary(lm)
```

```
##
## Call:
## lm(formula = Employed ~ ., data = ds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41011 -0.15767 -0.02816  0.10155  0.45539
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
```

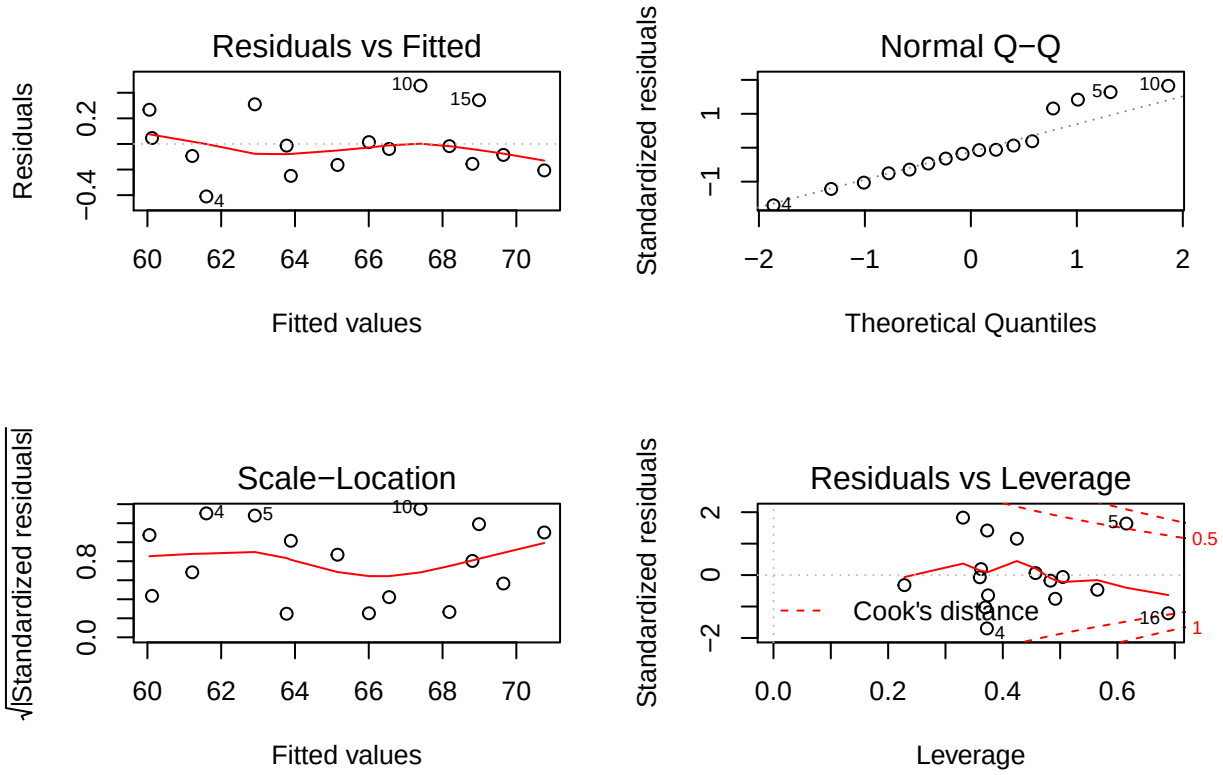
```
## (Intercept)  -3.482e+03  8.904e+02  -3.911 0.003560 **
## GNP          -3.582e-02  3.349e-02  -1.070 0.312681
## GNP.deflator  1.506e-02  8.492e-02   0.177 0.863141
## Unemployed   -2.020e-02  4.884e-03  -4.136 0.002535 **
## Armed.Forces -1.033e-02  2.143e-03  -4.822 0.000944 ***
## Population   -5.110e-02  2.261e-01  -0.226 0.826212
## Year          1.829e+00  4.555e-01   4.016 0.003037 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3049 on 9 degrees of freedom
## Multiple R-squared:  0.9955, Adjusted R-squared:  0.9925
## F-statistic: 330.3 on 6 and 9 DF,  p-value: 4.984e-10
```

```
anova(lm)
```

```
## Analysis of Variance Table
##
## Response: Employed
##              Df  Sum Sq Mean Sq   F value    Pr(>F)
## GNP           1 178.973 178.973 1925.7626 8.277e-12 ***
## GNP.deflator  1   0.212   0.212    2.2805 0.1652873
## Unemployed    1   2.264   2.264   24.3605 0.0008071 ***
## Armed.Forces  1   0.876   0.876    9.4301 0.0133357 *
## Population    1   0.349   0.349    3.7509 0.0847552 .
## Year          1   1.499   1.499   16.1274 0.0030368 **
## Residuals     9   0.836   0.093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

diagnostics

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



Co-relation matrix of regressors

```
regressors_matrix=as.matrix(ds[, -7])
regressors_matrix
```

```
##           GNP GNP.deflator Unemployed Armed.Forces Population Year
##  [1,] 234.289         83.0      235.6        159.0    107.608 1947
##  [2,] 259.426         88.5      232.5        145.6    108.632 1948
##  [3,] 258.054         88.2      368.2        161.6    109.773 1949
##  [4,] 284.599         89.5      335.1        165.0    110.929 1950
##  [5,] 328.975         96.2      209.9        309.9    112.075 1951
##  [6,] 346.999         98.1      193.2        359.4    113.270 1952
##  [7,] 365.385         99.0      187.0        354.7    115.094 1953
##  [8,] 363.112        100.0      357.8        335.0    116.219 1954
##  [9,] 397.469        101.2      290.4        304.8    117.388 1955
## [10,] 419.180        104.6      282.2        285.7    118.734 1956
## [11,] 442.769        108.4      293.6        279.8    120.445 1957
## [12,] 444.546        110.8      468.1        263.7    121.950 1958
## [13,] 482.704        112.6      381.3        255.2    123.366 1959
## [14,] 502.601        114.2      393.1        251.4    125.368 1960
## [15,] 518.173        115.7      480.6        257.2    127.852 1961
## [16,] 554.894        116.9      400.7        282.7    130.081 1962
```

```r
cor(regressors_matrix)
```

```
##                     GNP GNP.deflator Unemployed Armed.Forces Population
## GNP          1.0000000    0.9915892  0.6042609    0.4464368  0.9910901
## GNP.deflator 0.9915892    1.0000000  0.6206334    0.4647442  0.9791634
## Unemployed   0.6042609    0.6206334  1.0000000   -0.1774206  0.6865515
## Armed.Forces 0.4464368    0.4647442 -0.1774206    1.0000000  0.3644163
## Population   0.9910901    0.9791634  0.6865515    0.3644163  1.0000000
## Year         0.9952735    0.9911492  0.6682566    0.4172451  0.9939528
##                   Year
## GNP          0.9952735
## GNP.deflator 0.9911492
## Unemployed   0.6682566
## Armed.Forces 0.4172451
## Population   0.9939528
## Year         1.0000000
```

eigen values

```r
eigns=eigen(t(regressors_matrix)%*%regressors_matrix)
eigns
```

```
## eigen() decomposition
## $values
## [1] 6.665299e+07 2.090730e+05 1.053550e+05 1.803976e+04 2.455730e+01
## [6] 2.015117e+00
##
## $vectors
##             [,1]        [,2]        [,3]        [,4]        [,5]
## [1,] -0.19075418  0.72496814 -0.34330489  0.55402997  0.07487553
## [2,] -0.04990131  0.06979071 -0.03416853  0.04265870 -0.95653127
## [3,] -0.15702286  0.62152746  0.56371985 -0.52067703  0.00716578
## [4,] -0.12796016  0.10434859 -0.74630465 -0.64468394  0.01222896
## [5,] -0.05758090  0.03841364 -0.01095845  0.03583083  0.28108541
## [6,] -0.95748481 -0.26625145  0.07812474  0.05679111  0.01522131
##               [,6]
## [1,]  0.0872940138
## [2,] -0.2733126381
## [3,]  0.0105568115
## [4,] -0.0001122542
## [5,] -0.9564496276
## [6,]  0.0526555591
```

condition number

```r
cond_number = sqrt(eigns$values[1]/eigns$values)
cond_number
```

```
## [1]    1.00000   17.85504   25.15256   60.78472 1647.47771 5751.21560
```

estimating the VIF for one regressor (example)

```
regressors_matrix[,1]
```

```
##  [1] 234.289 259.426 258.054 284.599 328.975 346.999 365.385 363.112
##  [9] 397.469 419.180 442.769 444.546 482.704 502.601 518.173 554.894
```

```
regressors_matrix[ , -1]
```

```
##       GNP.deflator Unemployed Armed.Forces Population Year
##  [1,]         83.0      235.6        159.0    107.608 1947
##  [2,]         88.5      232.5        145.6    108.632 1948
##  [3,]         88.2      368.2        161.6    109.773 1949
##  [4,]         89.5      335.1        165.0    110.929 1950
##  [5,]         96.2      209.9        309.9    112.075 1951
##  [6,]         98.1      193.2        359.4    113.270 1952
##  [7,]         99.0      187.0        354.7    115.094 1953
##  [8,]        100.0      357.8        335.0    116.219 1954
##  [9,]        101.2      290.4        304.8    117.388 1955
## [10,]        104.6      282.2        285.7    118.734 1956
## [11,]        108.4      293.6        279.8    120.445 1957
## [12,]        110.8      468.1        263.7    121.950 1958
## [13,]        112.6      381.3        255.2    123.366 1959
## [14,]        114.2      393.1        251.4    125.368 1960
## [15,]        115.7      480.6        257.2    127.852 1961
## [16,]        116.9      400.7        282.7    130.081 1962
```

```
model1=lm(regressors_matrix[,1]~regressors_matrix[ , -1])
summary(model1)
```

```
##
## Call:
## lm(formula = regressors_matrix[, 1] ~ regressors_matrix[, -1])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8685 -1.4809 -0.3626  1.5030  4.9323
##
## Coefficients:
##                                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)                       -2.170e+04  4.859e+03  -4.466 0.001205
## regressors_matrix[, -1]GNP.deflator  1.647e+00  6.097e-01   2.701 0.022289
## regressors_matrix[, -1]Unemployed   -1.379e-01  1.500e-02  -9.192 3.42e-06
## regressors_matrix[, -1]Armed.Forces -2.998e-02  1.787e-02  -1.677 0.124388
## regressors_matrix[, -1]Population    5.624e+00  1.180e+00   4.765 0.000763
## regressors_matrix[, -1]Year          1.090e+01  2.571e+00   4.241 0.001713
##
## (Intercept)                       **
## regressors_matrix[, -1]GNP.deflator *
## regressors_matrix[, -1]Unemployed   ***
## regressors_matrix[, -1]Armed.Forces
```

```
## regressors_matrix[, -1]Population   ***
## regressors_matrix[, -1]Year         **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.878 on 10 degrees of freedom
## Multiple R-squared:  0.9994, Adjusted R-squared:  0.9992
## F-statistic:  3575 on 5 and 10 DF,  p-value: 6.405e-16
```

```
r_sqr_1 = summary(model1)$r.squared
r_sqr_1
```

```
## [1] 0.9994409
```

```
vif_1 = 1/(1-r_sqr_1)
vif_1
```

```
## [1] 1788.513
```

all vifs in one go

```
vif(lm)
```

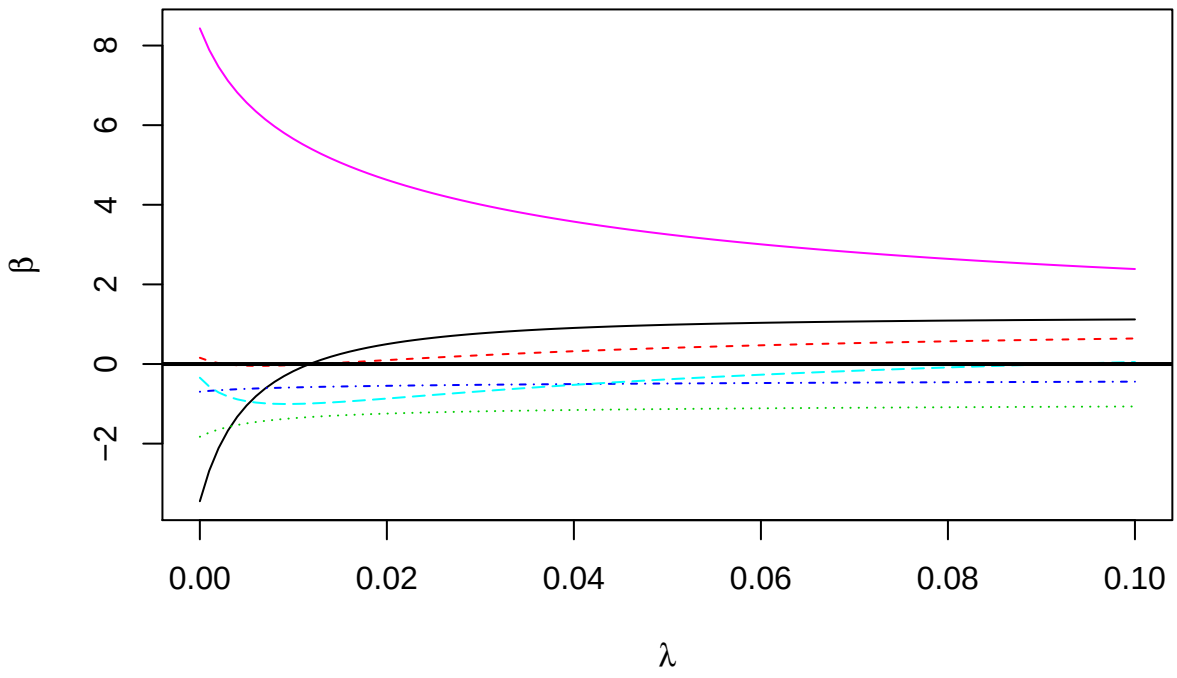
```
##          GNP GNP.deflator   Unemployed Armed.Forces   Population
##   1788.51348    135.53244     33.61889      3.58893    399.15102
##         Year
##    758.98060
```

```
ridgegrd = lm.ridge(Employed~., data=ds,lambda=seq(0,0.1, 0.001))
#ridgegrd
```

```
names(ridgegrd)
```

```
## [1] "coef"   "scales" "Inter"  "lambda" "ym"     "xm"     "GCV"    "kHKB"
## [9] "kLW"
```

```
matplot(ridgegrd$lambda, t(ridgegrd$coef), type="l",xlab=expression(lambda), ylab=expression(hat(beta)))
abline(h=0, lwd=2)
```



choosing the lambda

Hoerl and Kennard

```
ridgegrd$kHKB
```

```
## [1] 0.004275357
```

Lawless and Wang

```
ridgegrd$kLW
```

```
## [1] 0.03229531
```

Generalized Cross Validation

```
min(ridgegrd$GCV)
```

```
## [1] 0.008054062
```

```
which.min(ridgegrd$GCV)
```

```
## 0.003
##     4
```

estimating the coefficents now

```
ridgegrd$coef[, ridgegrd$lambda==0.03]
```

```
##          GNP GNP.deflator   Unemployed Armed.Forces   Population
##    0.7693585    0.2200496   -1.1894102   -0.5223393   -0.6861816
##         Year
##    4.0064269
```

```
lm$coefficients[-1]
```

```
##          GNP GNP.deflator   Unemployed Armed.Forces   Population
##  -0.03581918   0.01506187  -0.02020230  -0.01033227  -0.05110411
##         Year
##   1.82915146
```