ProjectTeam5

ProjectTeam5

3/15/2021

step 1 read and modify dataset

```
who.data       <- read.csv("Life Expectancy Data.csv"  )
who.data.2015 <- who.data[who.data$Year==2015,]
head(who.data.2015)
```

```
##                   Country Year    Status Life.expectancy Adult.Mortality
## 1             Afghanistan 2015 Developing            65.0             263
## 17                Albania 2015 Developing            77.8              74
## 33                Algeria 2015 Developing            75.6              19
## 49                 Angola 2015 Developing            52.4             335
## 65 Antigua and Barbuda 2015 Developing            76.4              13
## 81              Argentina 2015 Developing            76.3             116
##    infant.deaths Alcohol percentage.expenditure Hepatitis.B Measles  BMI
## 1             62    0.01                71.27962          65    1154 19.1
## 17             0    4.60               364.97523          99       0 58.0
## 33            21      NA                 0.00000          95      63 59.5
## 49            66      NA                 0.00000          64     118 23.3
## 65             0      NA                 0.00000          99       0 47.7
## 81             8      NA                 0.00000          94       0 62.8
##    under.five.deaths Polio Total.expenditure Diphtheria HIV.AIDS        GDP
## 1                 83     6              8.16          65      0.1   584.2592
## 17                 0    99              6.00          99      0.1  3954.2278
## 33                24    95                NA          95      0.1  4132.7629
## 49                98     7                NA          64      1.9  3695.7937
## 65                 0    86                NA          99      0.2 13566.9541
## 81                 9    93                NA          94      0.1 13467.1236
##    Population thinness..1.19.years thinness.5.9.years
## 1    33736494                 17.2               17.3
## 17      28873                  1.2                1.3
## 33   39871528                  6.0                5.8
## 49    2785935                  8.3                8.2
## 65         NA                  3.3                3.3
## 81   43417765                  1.0                0.9
##    Income.composition.of.resources Schooling
## 1                            0.479      10.1
## 17                           0.762      14.2
## 33                           0.743      14.4
## 49                           0.531      11.4
## 65                           0.784      13.9
## 81                           0.826      17.3
```

```r
#simplify the field's name for Rcode
names(who.data.2015) <-
  c("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O","P","Q","R","S","T","U","V")
nrow(who.data.2015)
```

```
## [1] 183
```

```r
#list the field name and it's letter
who.name=
  c("Country","Year","Status","Life expectancy","Adult Mortality","infant deaths","Alcohol",
    "percentage expenditure","Hepatitis B","Measles","BMI","under-five deaths","Polio",
    "Total expenditure","Diphtheria","HIV/AIDS","GDP","Population","thinness  1-19 years",
    "thinness 5-9 years","Income composition of resources","Schooling")
who.letter=
  c("A","B","C","D","E","F","G","H","I","J","K","L","M","N","O","P","Q","R","S","T","U","V")

library(data.table)
```

```
## Warning: package 'data.table' was built under R version 4.0.4
```

```r
data.table(who.name,who.letter)
```

```
##                          who.name who.letter
##  1:                        Country          A
##  2:                           Year          B
##  3:                         Status          C
##  4:                Life expectancy          D
##  5:                Adult Mortality          E
##  6:                  infant deaths          F
##  7:                        Alcohol          G
##  8:         percentage expenditure          H
##  9:                    Hepatitis B          I
## 10:                        Measles          J
## 11:                            BMI          K
## 12:              under-five deaths          L
## 13:                          Polio          M
## 14:              Total expenditure          N
## 15:                      Diphtheria          O
## 16:                       HIV/AIDS          P
## 17:                            GDP          Q
## 18:                     Population          R
## 19:           thinness  1-19 years          S
## 20:             thinness 5-9 years          T
## 21: Income composition of resources          U
## 22:                      Schooling          V
##                          who.name who.letter
```

step 2 using backward,forword,both,regsubsets, to determine best predictors; drop A,B,G,H,N.(empty too much);

```r
library(olsrr)
```

```
## Warning: package 'olsrr' was built under R version 4.0.4
```

```
##
```

```
## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##
##      rivers

who.fullmodel <- lm(D~factor(C)+E+F+I+J+K+L+M+O+P+Q+R+S+T+U+V,
                    data=who.data.2015)
summary(who.fullmodel)  #adj.r.squared 0.8865,E,F,I,L,P,U are predictors (I p-value 0.0577 )
```

```
##
## Call:
## lm(formula = D ~ factor(C) + E + F + I + J + K + L + M + O +
##     P + Q + R + S + T + U + V, data = who.data.2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3416 -1.4379  0.0359  1.5459  7.9402
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          5.040e+01  2.511e+00  20.074  < 2e-16 ***
## factor(C)Developing -3.424e-01  8.299e-01  -0.413   0.6807
## E                   -2.092e-02  3.607e-03  -5.799 6.15e-08 ***
## F                    6.601e-02  3.285e-02   2.010   0.0469 *
## I                    4.333e-02  2.260e-02   1.917   0.0577 .
## J                   -5.119e-05  5.729e-05  -0.893   0.3735
## K                   -8.580e-03  1.550e-02  -0.554   0.5809
## L                   -4.783e-02  2.354e-02  -2.032   0.0445 *
## M                    1.147e-02  1.267e-02   0.905   0.3676
## O                   -1.106e-02  2.630e-02  -0.420   0.6750
## P                   -4.847e-01  2.239e-01  -2.165   0.0325 *
## Q                    5.064e-06  3.003e-05   0.169   0.8664
## R                   -1.010e-09  9.564e-09  -0.106   0.9161
## S                   -1.228e-01  2.338e-01  -0.525   0.6004
## T                   -1.735e-02  2.287e-01  -0.076   0.9396
## U                    3.325e+01  4.981e+00   6.676 9.59e-10 ***
## V                   -4.796e-02  2.402e-01  -0.200   0.8421
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.695 on 113 degrees of freedom
##   (53 observations deleted due to missingness)
## Multiple R-squared:  0.9005, Adjusted R-squared:  0.8865
## F-statistic: 63.95 on 16 and 113 DF,  p-value: < 2.2e-16
```

```
#use three methods to get best predictors
who.backwardmodel <- ols_step_backward_p(who.fullmodel,prem = 0.05,details = TRUE)
```

```
## Backward Elimination Method
## ---------------------------
##
## Candidate Terms:
##
## 1 . factor(C)
## 2 . E
```

```
## 3 . F
## 4 . I
## 5 . J
## 6 . K
## 7 . L
## 8 . M
## 9 . O
## 10 . P
## 11 . Q
## 12 . R
## 13 . S
## 14 . T
## 15 . U
## 16 . V
##
## We are eliminating variables based on p value...
##
## - T
##
## Backward Elimination: Step 1
##
##   Variable T Removed
##
##                        Model Summary
## -----------------------------------------------------------------
## R                        0.949      RMSE              2.683
## R-Squared                0.901      Coef. Var         3.793
## Adj. R-Squared           0.887      MSE               7.198
## Pred R-Squared           0.856      MAE               1.948
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## --------------------------------------------------------------------------
##              Sum of
##              Squares      DF    Mean Square      F        Sig.
## --------------------------------------------------------------------------
## Regression   7429.971     15       495.331    68.811     0.0000
## Residual      820.625    114         7.198
## Total        8250.596    129
## --------------------------------------------------------------------------
##
##                          Parameter Estimates
## -----------------------------------------------------------------------------------------
##              model      Beta    Std. Error   Std. Beta     t       Sig     lower    upper
## -----------------------------------------------------------------------------------------
##        (Intercept)    50.392       2.495                20.198    0.000   45.450   55.334
## factor(C)Developing   -0.345       0.826      -0.015    -0.417    0.677   -1.980    1.291
##                  E    -0.021       0.004      -0.261    -5.880    0.000   -0.028   -0.014
##                  F     0.066       0.032       0.792     2.025    0.045    0.001    0.130
##                  I     0.043       0.023       0.135     1.926    0.057   -0.001    0.088
##                  J     0.000       0.000      -0.058    -0.895    0.373    0.000    0.000
```

4

```
##                        K      -0.008         0.015        -0.022    -0.552    0.582    -0.038      0.022
##                        L      -0.048         0.023        -0.738    -2.045    0.043    -0.094     -0.001
##                        M       0.011         0.013         0.036     0.905    0.367    -0.014      0.036
##                        O      -0.011         0.026        -0.031    -0.416    0.678    -0.062      0.041
##                        P      -0.482         0.219        -0.092    -2.200    0.030    -0.915     -0.048
##                        Q       0.000         0.000         0.006     0.171    0.865     0.000      0.000
##                        R       0.000         0.000        -0.004    -0.104    0.918     0.000      0.000
##                        S      -0.139         0.084        -0.076    -1.657    0.100    -0.306      0.027
##                        U      33.245         4.958         0.629     6.705    0.000    23.423     43.068
##                        V      -0.049         0.239        -0.017    -0.205    0.838    -0.522      0.424
## ----------------------------------------------------------------------------------------------------
##
##
## - R
##
## Backward Elimination: Step 2
##
##   Variable R Removed
##
##                         Model Summary
## --------------------------------------------------------------
## R                          0.950       RMSE              2.591
## R-Squared                  0.903       Coef. Var         3.632
## Adj. R-Squared             0.892       MSE               6.713
## Pred R-Squared             0.866       MAE               1.895
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## --------------------------------------------------------------------------
##                Sum of
##               Squares        DF     Mean Square      F         Sig.
## --------------------------------------------------------------------------
## Regression   7990.229        14         570.731    85.018     0.0000
## Residual      859.269       128           6.713
## Total        8849.498       142
## --------------------------------------------------------------------------
##
##                            Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##                model      Beta     Std. Error    Std. Beta      t        Sig      lower      upper
## ----------------------------------------------------------------------------------------------------
##          (Intercept)    51.374        2.300                   22.336    0.000    46.823     55.925
## factor(C)Developing     -0.670        0.751        -0.030     -0.892    0.374    -2.157      0.817
##                    E    -0.021        0.003        -0.255     -6.044    0.000    -0.027     -0.014
##                    F     0.072        0.030         0.841      2.409    0.017     0.013      0.131
##                    I     0.040        0.021         0.122      1.866    0.064    -0.002      0.082
##                    J     0.000        0.000        -0.061     -1.006    0.316     0.000      0.000
##                    K    -0.011        0.014        -0.030     -0.791    0.430    -0.039      0.017
##                    L    -0.052        0.021        -0.779     -2.471    0.015    -0.093     -0.010
##                    M     0.012        0.012         0.037      1.019    0.310    -0.011      0.035
##                    O    -0.008        0.025        -0.024     -0.338    0.736    -0.057      0.041
```

```
##                     P    -0.507         0.210       -0.094  -2.417  0.017   -0.922  -0.092
##                     Q     0.000         0.000       -0.004  -0.117  0.907    0.000   0.000
##                     S    -0.168         0.078       -0.090  -2.150  0.033   -0.322  -0.013
##                     U    31.946         4.487        0.613   7.119  0.000   23.067  40.825
##                     V    -0.019         0.210       -0.007  -0.089  0.929   -0.434   0.397
## ----------------------------------------------------------------------------------------------
##
##
## - V
##
## Backward Elimination: Step 3
##
##   Variable V Removed
##
##                      Model Summary
## --------------------------------------------------------------
## R                        0.950      RMSE                2.581
## R-Squared                0.903      Coef. Var           3.618
## Adj. R-Squared           0.893      MSE                 6.661
## Pred R-Squared           0.868      MAE                 1.897
## --------------------------------------------------------------
##   RMSE: Root Mean Square Error
##   MSE: Mean Square Error
##   MAE: Mean Absolute Error
##
##                           ANOVA
## ---------------------------------------------------------------------
##              Sum of
##              Squares        DF     Mean Square      F        Sig.
## ---------------------------------------------------------------------
## Regression   7990.176       13         614.629    92.267   0.0000
## Residual      859.323      129           6.661
## Total        8849.498      142
## ---------------------------------------------------------------------
##
##
##                         Parameter Estimates
## ----------------------------------------------------------------------------------------------
##               model     Beta    Std. Error    Std. Beta      t      Sig     lower   upper
## ----------------------------------------------------------------------------------------------
##        (Intercept)    51.346       2.271                   22.611  0.000   46.854  55.839
## factor(C)Developing   -0.657       0.735       -0.030      -0.894  0.373   -2.111   0.797
##                  E    -0.021       0.003       -0.256      -6.220  0.000   -0.027  -0.014
##                  F     0.072       0.030        0.842       2.424  0.017    0.013   0.131
##                  I     0.040       0.021        0.122       1.875  0.063   -0.002   0.081
##                  J     0.000       0.000       -0.061      -1.025  0.307    0.000   0.000
##                  K    -0.011       0.014       -0.030      -0.807  0.421   -0.039   0.016
##                  L    -0.052       0.021       -0.780      -2.483  0.014   -0.093  -0.011
##                  M     0.012       0.011        0.037       1.034  0.303   -0.011   0.035
##                  O    -0.008       0.025       -0.024      -0.335  0.738   -0.057   0.040
##                  P    -0.507       0.209       -0.094      -2.426  0.017   -0.920  -0.093
##                  Q     0.000       0.000       -0.004      -0.113  0.910    0.000   0.000
##                  S    -0.168       0.078       -0.090      -2.157  0.033   -0.321  -0.014
##                  U    31.636       2.832        0.607      11.170  0.000   26.032  37.239
## ----------------------------------------------------------------------------------------------
```

```
## 
## 
## - Q
## 
## Backward Elimination: Step 4
## 
##   Variable Q Removed
## 
##                         Model Summary
## -----------------------------------------------------------
## R                       0.945      RMSE               2.601
## R-Squared               0.893      Coef. Var          3.647
## Adj. R-Squared          0.884      MSE                6.763
## Pred R-Squared          0.854      MAE                1.921
## -----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                             ANOVA
## --------------------------------------------------------------------
##               Sum of
##               Squares        DF     Mean Square      F        Sig.
## --------------------------------------------------------------------
## Regression    8441.275       12         703.440    104.013   0.0000
## Residual      1014.454      150           6.763
## Total         9455.728      162
## --------------------------------------------------------------------
## 
## 
##                            Parameter Estimates
## ----------------------------------------------------------------------------------
##            model       Beta    Std. Error    Std. Beta      t       Sig    lower    upper
## ----------------------------------------------------------------------------------
##       (Intercept)     51.135     2.174                    23.518   0.000   46.838   55.431
## factor(C)Developing   -0.485     0.708        -0.022      -0.685    0.494   -1.884    0.913
##                 E     -0.018     0.003        -0.229      -5.732    0.000   -0.025   -0.012
##                 F      0.070     0.029         0.798       2.424    0.017    0.013    0.128
##                 I      0.037     0.021         0.118       1.773    0.078   -0.004    0.079
##                 J      0.000     0.000        -0.048      -0.865    0.389    0.000    0.000
##                 K     -0.006     0.013        -0.016      -0.441    0.660   -0.032    0.020
##                 L     -0.052     0.021        -0.761      -2.532    0.012   -0.093   -0.011
##                 M      0.011     0.011         0.036       0.994    0.322   -0.011    0.033
##                 O     -0.010     0.024        -0.028      -0.394    0.694   -0.058    0.039
##                 P     -0.643     0.200        -0.117      -3.210    0.002   -1.039   -0.247
##                 S     -0.125     0.071        -0.068      -1.748    0.083   -0.265    0.016
##                 U     31.459     2.562         0.609      12.281    0.000   26.398   36.520
## ----------------------------------------------------------------------------------
## 
## 
## - O
## 
## Backward Elimination: Step 5
## 
##   Variable O Removed
```

```
##
##                         Model Summary
## ---------------------------------------------------------------
## R                         0.945       RMSE              2.593
## R-Squared                 0.893       Coef. Var         3.637
## Adj. R-Squared            0.885       MSE               6.725
## Pred R-Squared            0.857       MAE               1.916
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## -------------------------------------------------------------------
##              Sum of
##              Squares       DF    Mean Square      F        Sig.
## -------------------------------------------------------------------
## Regression  8440.227       11       767.293    114.093    0.0000
## Residual    1015.502      151         6.725
## Total       9455.728      162
## -------------------------------------------------------------------
##
##
##                           Parameter Estimates
## --------------------------------------------------------------------------------------
##              model      Beta    Std. Error    Std. Beta     t       Sig     lower     upper
## --------------------------------------------------------------------------------------
##       (Intercept)     51.085      2.164                    23.601   0.000   46.808    55.361
## factor(C)Developing   -0.461      0.703        -0.021      -0.655   0.513   -1.850     0.928
##                   E   -0.018      0.003        -0.229      -5.760   0.000   -0.025    -0.012
##                   F    0.071      0.029         0.802       2.444   0.016    0.014     0.128
##                   I    0.030      0.011         0.095       2.789   0.006    0.009     0.052
##                   J    0.000      0.000        -0.049      -0.885   0.378    0.000     0.000
##                   K   -0.005      0.013        -0.014      -0.398   0.691   -0.031     0.020
##                   L   -0.052      0.021        -0.763      -2.547   0.012   -0.093    -0.012
##                   M    0.010      0.011         0.033       0.928   0.355   -0.011     0.031
##                   P   -0.636      0.199        -0.116      -3.196   0.002   -1.029    -0.243
##                   S   -0.127      0.071        -0.070      -1.790   0.075   -0.267     0.013
##                   U   31.278      2.513         0.606      12.447   0.000   26.313    36.243
## --------------------------------------------------------------------------------------
##
##
## - K
##
## Backward Elimination: Step 6
##
##  Variable K Removed
##
##                         Model Summary
## ---------------------------------------------------------------
## R                         0.945       RMSE              2.586
## R-Squared                 0.892       Coef. Var         3.627
## Adj. R-Squared            0.885       MSE               6.688
## Pred R-Squared            0.858       MAE               1.916
## ---------------------------------------------------------------
```

```
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------
##               Sum of
##               Squares        DF     Mean Square        F       Sig.
## ---------------------------------------------------------------------
## Regression    8439.162       10        843.916     126.185    0.0000
## Residual      1016.566      152          6.688
## Total         9455.728      162
## ---------------------------------------------------------------------
##
##
##                          Parameter Estimates
## -------------------------------------------------------------------------------------
##               model       Beta    Std. Error   Std. Beta      t       Sig    lower     upper
## -------------------------------------------------------------------------------------
##         (Intercept)      51.130     2.155                   23.722   0.000   46.872   55.389
## factor(C)Developing      -0.485     0.698        -0.022     -0.695   0.488   -1.865    0.894
##                   E      -0.018     0.003        -0.230     -5.802   0.000   -0.025   -0.012
##                   F       0.070     0.029         0.795      2.432   0.016    0.013    0.127
##                   I       0.030     0.011         0.095      2.780   0.006    0.009    0.051
##                   J       0.000     0.000        -0.047     -0.861   0.390    0.000    0.000
##                   L      -0.052     0.020        -0.758     -2.538   0.012   -0.092   -0.012
##                   M       0.010     0.011         0.034      0.979   0.329   -0.011    0.032
##                   P      -0.632     0.198        -0.115     -3.191   0.002   -1.024   -0.241
##                   S      -0.120     0.069        -0.066     -1.750   0.082   -0.256    0.015
##                   U      30.861     2.277         0.598     13.551   0.000   26.361   35.360
## -------------------------------------------------------------------------------------
##
##
## - factor(C)
##
## Backward Elimination: Step 7
##
##  Variable factor(C) Removed
##
##                      Model Summary
## -----------------------------------------------------------------
## R                       0.945     RMSE                2.582
## R-Squared               0.892     Coef. Var           3.621
## Adj. R-Squared          0.886     MSE                 6.665
## Pred R-Squared          0.862     MAE                 1.930
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------
##               Sum of
##               Squares        DF     Mean Square        F       Sig.
## ---------------------------------------------------------------------
```

```
## Regression      8435.933          9        937.326    140.627     0.0000
## Residual        1019.796        153          6.665
## Total           9455.728        162
## -----------------------------------------------------------------------
##
##
##                          Parameter Estimates
## ------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t       Sig     lower     upper
## ------------------------------------------------------------------------------
## (Intercept)    50.372       1.855                    27.149    0.000    46.706    54.037
##           E    -0.018       0.003       -0.231       -5.847    0.000    -0.025    -0.012
##           F     0.070       0.029        0.793        2.429    0.016     0.013     0.127
##           I     0.030       0.011        0.095        2.784    0.006     0.009     0.051
##           J     0.000       0.000       -0.049       -0.897    0.371     0.000     0.000
##           L    -0.052       0.020       -0.752       -2.526    0.013    -0.092    -0.011
##           M     0.010       0.011        0.034        0.975    0.331    -0.011     0.031
##           P    -0.621       0.197       -0.113       -3.150    0.002    -1.011    -0.232
##           S    -0.126       0.068       -0.069       -1.840    0.068    -0.260     0.009
##           U    31.405       2.135        0.608       14.713    0.000    27.188    35.622
## ------------------------------------------------------------------------------
##
##
## - J
##
## Backward Elimination: Step 8
##
##   Variable J Removed
##
##                       Model Summary
## --------------------------------------------------------------
## R                        0.944        RMSE                2.580
## R-Squared                0.892        Coef. Var           3.618
## Adj. R-Squared           0.886        MSE                 6.657
## Pred R-Squared           0.861        MAE                 1.924
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## -----------------------------------------------------------------------
##              Sum of
##              Squares         DF    Mean Square      F        Sig.
## -----------------------------------------------------------------------
## Regression   8430.573          8      1053.822    158.306    0.0000
## Residual     1025.156        154         6.657
## Total        9455.728        162
## -----------------------------------------------------------------------
##
##                          Parameter Estimates
## ------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t       Sig     lower     upper
## ------------------------------------------------------------------------------
## (Intercept)    50.247       1.849                    27.176    0.000    46.595    53.900
```

```
##              E    -0.018          0.003          -0.229    -5.812    0.000    -0.025    -0.012
##              F     0.054          0.023           0.615     2.373    0.019     0.009     0.099
##              I     0.029          0.011           0.091     2.707    0.008     0.008     0.050
##              L    -0.042          0.018          -0.618    -2.402    0.018    -0.077    -0.008
##              M     0.010          0.011           0.034     0.970    0.334    -0.011     0.031
##              P    -0.624          0.197          -0.114    -3.166    0.002    -1.013    -0.235
##              S    -0.110          0.066          -0.060    -1.668    0.097    -0.240     0.020
##              U    31.593          2.123           0.612    14.882    0.000    27.399    35.787
## -------------------------------------------------------------------------------------------
##
##
## - M
##
## Backward Elimination: Step 9
##
##   Variable M Removed
##
##                            Model Summary
## --------------------------------------------------------------
## R                       0.944       RMSE                 2.580
## R-Squared               0.891       Coef. Var            3.618
## Adj. R-Squared          0.886       MSE                  6.654
## Pred R-Squared          0.864       MAE                  1.915
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------
##                   Sum of
##                  Squares       DF    Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression      8424.314        7       1203.473    180.857    0.0000
## Residual        1031.415      155          6.654
## Total           9455.728      162
## --------------------------------------------------------------------
##
##                           Parameter Estimates
## ------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t       Sig    lower    upper
## ------------------------------------------------------------------------------
## (Intercept)    50.455        1.836                  27.478    0.000    46.828    54.082
##           E    -0.018        0.003       -0.231     -5.878    0.000    -0.025    -0.012
##           F     0.055        0.023        0.618      2.384    0.018     0.009     0.100
##           I     0.034        0.009        0.107      3.609    0.000     0.015     0.052
##           L    -0.043        0.018       -0.622     -2.417    0.017    -0.077    -0.008
##           P    -0.636        0.197       -0.116     -3.232    0.002    -1.024    -0.247
##           S    -0.110        0.066       -0.060     -1.668    0.097    -0.240     0.020
##           U    31.991        2.083        0.620     15.361    0.000    27.877    36.104
## ------------------------------------------------------------------------------
##
##
## - S
```

11

```
##
## Backward Elimination: Step 10
##
##   Variable S Removed
##
##                        Model Summary
## ----------------------------------------------------------------
## R                     0.944         RMSE              2.581
## R-Squared             0.891         Coef. Var         3.626
## Adj. R-Squared        0.887         MSE               6.662
## Pred R-Squared        0.869         MAE               1.919
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## --------------------------------------------------------------------
##              Sum of
##              Squares       DF     Mean Square      F         Sig.
## --------------------------------------------------------------------
## Regression   8648.334       6      1441.389     216.351    0.0000
## Residual     1052.639      158        6.662
## Total        9700.973      164
## --------------------------------------------------------------------
##
##                        Parameter Estimates
## ------------------------------------------------------------------------------
##      model       Beta    Std. Error    Std. Beta      t       Sig     lower    upper
## ------------------------------------------------------------------------------
## (Intercept)    49.089      1.610                    30.489    0.000   45.909   52.269
##         E      -0.018      0.003        -0.223      -5.712    0.000   -0.024   -0.012
##         F       0.046      0.022         0.512       2.058    0.041    0.002    0.090
##         I       0.031      0.009         0.098       3.367    0.001    0.013    0.049
##         L      -0.038      0.017        -0.542      -2.169    0.032   -0.072   -0.003
##         P      -0.643      0.196        -0.117      -3.279    0.001   -1.030   -0.256
##         U      33.530      1.837         0.651      18.249    0.000   29.901   37.159
## ------------------------------------------------------------------------------
##
##
##
## No more variables satisfy the condition of p value = 0.05
##
##
## Variables Removed:
##
## - T
## - R
## - V
## - Q
## - O
## - K
## - factor(C)
## - J
```

```
## - M
## - S
##
##
## Final Model Output
## ------------------
##
##
##                         Model Summary
## --------------------------------------------------------------
## R                        0.944       RMSE                2.581
## R-Squared                0.891       Coef. Var           3.626
## Adj. R-Squared           0.887       MSE                 6.662
## Pred R-Squared           0.869       MAE                 1.919
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##
##                              ANOVA
## --------------------------------------------------------------------
##                Sum of
##               Squares        DF    Mean Square      F          Sig.
## --------------------------------------------------------------------
## Regression    8648.334        6       1441.389   216.351      0.0000
## Residual      1052.639      158          6.662
## Total         9700.973      164
## --------------------------------------------------------------------
##
##
##                            Parameter Estimates
## ----------------------------------------------------------------------------------
##        model       Beta    Std. Error    Std. Beta       t       Sig      lower      upper
## ----------------------------------------------------------------------------------
## (Intercept)      49.089       1.610                    30.489    0.000    45.909     52.269
##          E       -0.018       0.003       -0.223       -5.712    0.000    -0.024     -0.012
##          F        0.046       0.022        0.512        2.058    0.041     0.002      0.090
##          I        0.031       0.009        0.098        3.367    0.001     0.013      0.049
##          L       -0.038       0.017       -0.542       -2.169    0.032    -0.072     -0.003
##          P       -0.643       0.196       -0.117       -3.279    0.001    -1.030     -0.256
##          U       33.530       1.837        0.651       18.249    0.000    29.901     37.159
## ----------------------------------------------------------------------------------
```

```r
summary(who.backwardmodel$model) #adj.r.squared 0.8874,E,F,I,L,P,U left
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.611 -1.417 -0.099  1.696  8.387
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.088788   1.610056  30.489  < 2e-16 ***
```

```
## E              -0.017843   0.003124  -5.712 5.41e-08 ***
## F               0.045692   0.022204   2.058 0.041245 *
## I               0.031078   0.009230   3.367 0.000954 ***
## L              -0.037638   0.017355  -2.169 0.031599 *
## P              -0.642729   0.196034  -3.279 0.001283 **
## U              33.529872   1.837307  18.249  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.581 on 158 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8915, Adjusted R-squared:  0.8874
## F-statistic: 216.4 on 6 and 158 DF,  p-value: < 2.2e-16
```

```r
who.forwardmodel <- ols_step_forward_p(who.fullmodel,penter=0.05,details=TRUE)
```

```
## Forward Selection Method
## ---------------------------
##
## Candidate Terms:
##
## 1. factor(C)
## 2. E
## 3. F
## 4. I
## 5. J
## 6. K
## 7. L
## 8. M
## 9. O
## 10. P
## 11. Q
## 12. R
## 13. S
## 14. T
## 15. U
## 16. V
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - U
##
##                        Model Summary
## --------------------------------------------------------------
## R                      0.907        RMSE                 3.345
## R-Squared              0.823        Coef. Var            4.664
## Adj. R-Squared         0.822        MSE                 11.186
## Pred R-Squared         0.819        MAE                  2.508
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
```

```
## 
##                                 ANOVA                                 
## -----------------------------------------------------------------
##                  Sum of                                              
##                 Squares       DF      Mean Square       F         Sig.
## -----------------------------------------------------------------
## Regression      8912.941        1        8912.941     796.792    0.0000
## Residual        1912.811      171          11.186                      
## Total          10825.752      172                                     
## -----------------------------------------------------------------
## 
## 
##                            Parameter Estimates                            
## --------------------------------------------------------------------------
##       model       Beta     Std. Error    Std. Beta       t        Sig      lower     upper
## --------------------------------------------------------------------------
## (Intercept)     39.254       1.178                     33.332     0.000    36.929    41.579
##           U     46.923       1.662        0.907        28.228     0.000    43.642    50.205
## --------------------------------------------------------------------------
## 
## 
## 
## Forward Selection: Step 2
## 
## - E
## 
##                         Model Summary                         
## ----------------------------------------------------------------
## R                      0.933        RMSE              2.874
## R-Squared              0.870        Coef. Var         4.007
## Adj. R-Squared         0.869        MSE               8.258
## Pred R-Squared         0.864        MAE               2.086
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
## 
##                                 ANOVA                                 
## -----------------------------------------------------------------
##                  Sum of                                              
##                 Squares       DF      Mean Square       F         Sig.
## -----------------------------------------------------------------
## Regression      9421.933        2        4710.967     570.49     0.0000
## Residual        1403.819      170           8.258                      
## Total          10825.752      172                                     
## -----------------------------------------------------------------
## 
##                            Parameter Estimates                            
## --------------------------------------------------------------------------
##       model       Beta     Std. Error    Std. Beta       t        Sig      lower     upper
## --------------------------------------------------------------------------
## (Intercept)     49.624       1.664                     29.824     0.000    46.340    52.909
##           U     37.144       1.895        0.718        19.600     0.000    33.403    40.885
##           E     -0.024       0.003       -0.288        -7.851     0.000    -0.030    -0.018
## --------------------------------------------------------------------------
```

```
##
##
##
## Forward Selection: Step 3
##
## - I
##
##                         Model Summary
## -----------------------------------------------------------
## R                       0.938       RMSE                2.698
## R-Squared               0.879       Coef. Var           3.790
## Adj. R-Squared          0.877       MSE                 7.277
## Pred R-Squared          0.870       MAE                 2.062
## -----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## ---------------------------------------------------------------------
##              Sum of
##              Squares        DF     Mean Square      F         Sig.
## ---------------------------------------------------------------------
## Regression   8529.407        3       2843.136    390.712     0.0000
## Residual     1171.566      161          7.277
## Total        9700.973      164
## ---------------------------------------------------------------------
##
##                              Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t       Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    47.908      1.649                     29.054    0.000    44.651    51.164
##         U      34.393      1.897        0.668        18.131    0.000    30.647    38.139
##         E      -0.024      0.003       -0.295        -8.255    0.000    -0.029    -0.018
##         I       0.042      0.009        0.134         4.630    0.000     0.024     0.061
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 4
##
## - P
##
##                         Model Summary
## -----------------------------------------------------------
## R                       0.942       RMSE                2.609
## R-Squared               0.888       Coef. Var           3.666
## Adj. R-Squared          0.885       MSE                 6.809
## Pred R-Squared          0.876       MAE                 1.955
## -----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
```

```
##
##                            ANOVA
## ------------------------------------------------------------------
##                 Sum of
##               Squares         DF     Mean Square        F        Sig.
## ------------------------------------------------------------------
## Regression     8611.466         4       2152.867     316.16     0.0000
## Residual       1089.507       160          6.809
## Total          9700.973       164
## ------------------------------------------------------------------
##
##
##                          Parameter Estimates
## -----------------------------------------------------------------------------
##        model       Beta    Std. Error    Std. Beta       t       Sig    lower    upper
## -----------------------------------------------------------------------------
## (Intercept)      48.601       1.608                    30.234    0.000   45.426   51.776
##           U      34.053       1.838        0.661       18.532    0.000   30.424   37.682
##           E      -0.018       0.003       -0.228       -5.790    0.000   -0.024   -0.012
##           I       0.033       0.009        0.103        3.499    0.001    0.014    0.051
##           P      -0.683       0.197       -0.124       -3.471    0.001   -1.071   -0.294
## -----------------------------------------------------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + U
## + E
## + I
## + P
##
##
## Final Model Output
## ------------------
##
##                          Model Summary
## ------------------------------------------------------------
## R                        0.942      RMSE               2.609
## R-Squared                0.888      Coef. Var          3.666
## Adj. R-Squared           0.885      MSE                6.809
## Pred R-Squared           0.876      MAE                1.955
## ------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## ------------------------------------------------------------------
##                 Sum of
##               Squares         DF     Mean Square        F        Sig.
## ------------------------------------------------------------------
## Regression     8611.466         4       2152.867     316.16     0.0000
```

```
## Residual       1089.507       160           6.809
## Total          9700.973       164
## -----------------------------------------------------------------------------
##
##                              Parameter Estimates
## -------------------------------------------------------------------------------------------
##       model       Beta     Std. Error     Std. Beta        t         Sig       lower      upper
## -------------------------------------------------------------------------------------------
## (Intercept)     48.601       1.608                       30.234     0.000      45.426     51.776
##           U     34.053       1.838         0.661         18.532     0.000      30.424     37.682
##           E     -0.018       0.003        -0.228         -5.790     0.000      -0.024     -0.012
##           I      0.033       0.009         0.103          3.499     0.001       0.014      0.051
##           P     -0.683       0.197        -0.124         -3.471     0.001      -1.071     -0.294
## -------------------------------------------------------------------------------------------
```

```r
summary(who.forwardmodel$model) #adj.r.squared 0.8849  #E,I,P,U left
```

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9589 -1.5002 -0.0549  1.7899  8.6629
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.600964   1.607515  30.234  < 2e-16 ***
## U           34.053141   1.837554  18.532  < 2e-16 ***
## E           -0.018245   0.003151  -5.790 3.61e-08 ***
## I            0.032575   0.009309   3.499 0.000604 ***
## P           -0.682684   0.196658  -3.471 0.000666 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.609 on 160 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8877, Adjusted R-squared:  0.8849
## F-statistic: 316.2 on 4 and 160 DF,  p-value: < 2.2e-16
```

```r
who.bothmodel   <- ols_step_both_p(who.fullmodel, pent=0.05, prem=0.05,details=TRUE)
```

```
## Stepwise Selection Method
## --------------------------
##
## Candidate Terms:
##
## 1. factor(C)
## 2. E
## 3. F
## 4. I
## 5. J
## 6. K
## 7. L
```

```
## 8. M
## 9. O
## 10. P
## 11. Q
## 12. R
## 13. S
## 14. T
## 15. U
## 16. V
##
## We are selecting variables based on p value...
##
##
## Stepwise Selection: Step 1
##
## - U added
##
##                          Model Summary
## ------------------------------------------------------------------
## R                        0.907      RMSE                  3.345
## R-Squared                0.823      Coef. Var             4.664
## Adj. R-Squared           0.822      MSE                  11.186
## Pred R-Squared           0.819      MAE                   2.508
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                              ANOVA
## --------------------------------------------------------------------------
##                Sum of
##               Squares        DF     Mean Square      F          Sig.
## --------------------------------------------------------------------------
## Regression    8912.941        1       8912.941     796.792     0.0000
## Residual      1912.811      171         11.186
## Total        10825.752      172
## --------------------------------------------------------------------------
##
##                          Parameter Estimates
## ---------------------------------------------------------------------------------
##      model      Beta    Std. Error    Std. Beta      t       Sig      lower    upper
## ---------------------------------------------------------------------------------
## (Intercept)    39.254      1.178                   33.332    0.000    36.929   41.579
##         U      46.923      1.662        0.907      28.228    0.000    43.642   50.205
## ---------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 2
##
## - E added
##
##                          Model Summary
## ------------------------------------------------------------------
```

19

```
## R                         0.933      RMSE                2.874
## R-Squared                 0.870      Coef. Var           4.007
## Adj. R-Squared            0.869      MSE                 8.258
## Pred R-Squared            0.864      MAE                 2.086
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## -------------------------------------------------------------------
##               Sum of
##               Squares        DF     Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression    9421.933        2       4710.967     570.49    0.0000
## Residual      1403.819      170          8.258
## Total        10825.752      172
## -------------------------------------------------------------------
##
##                        Parameter Estimates
## -----------------------------------------------------------------------------
##       model      Beta   Std. Error   Std. Beta      t        Sig     lower      upper
## -----------------------------------------------------------------------------
## (Intercept)    49.624      1.664                  29.824    0.000    46.340    52.909
##         U      37.144      1.895       0.718      19.600    0.000    33.403    40.885
##         E      -0.024      0.003      -0.288      -7.851    0.000    -0.030    -0.018
## -----------------------------------------------------------------------------
##
##
##
##                        Model Summary
## ----------------------------------------------------------------
## R                         0.933      RMSE                2.874
## R-Squared                 0.870      Coef. Var           4.007
## Adj. R-Squared            0.869      MSE                 8.258
## Pred R-Squared            0.864      MAE                 2.086
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## -------------------------------------------------------------------
##               Sum of
##               Squares        DF     Mean Square      F         Sig.
## -------------------------------------------------------------------
## Regression    9421.933        2       4710.967     570.49    0.0000
## Residual      1403.819      170          8.258
## Total        10825.752      172
## -------------------------------------------------------------------
##
##                        Parameter Estimates
## -----------------------------------------------------------------------------
##       model      Beta   Std. Error   Std. Beta      t        Sig     lower      upper
```

```
## ------------------------------------------------------------------------------
## (Intercept)    49.624         1.664                      29.824   0.000   46.340   52.909
##         U      37.144         1.895           0.718      19.600   0.000   33.403   40.885
##         E      -0.024         0.003          -0.288      -7.851   0.000   -0.030   -0.018
## ------------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 3
##
## - I added
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.938       RMSE              2.698
## R-Squared               0.879       Coef. Var         3.790
## Adj. R-Squared          0.877       MSE               7.277
## Pred R-Squared          0.870       MAE               2.062
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                             ANOVA
## --------------------------------------------------------------------
##              Sum of
##              Squares         DF    Mean Square       F         Sig.
## --------------------------------------------------------------------
## Regression   8529.407         3      2843.136    390.712     0.0000
## Residual     1171.566       161         7.277
## Total        9700.973       164
## --------------------------------------------------------------------
##
##                          Parameter Estimates
## ----------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig     lower    upper
## ----------------------------------------------------------------------------------
## (Intercept)    47.908       1.649                    29.054    0.000   44.651   51.164
##         U      34.393       1.897        0.668       18.131    0.000   30.647   38.139
##         E      -0.024       0.003       -0.295       -8.255    0.000   -0.029   -0.018
##         I       0.042       0.009        0.134        4.630    0.000    0.024    0.061
## ----------------------------------------------------------------------------------
##
##
##
##                         Model Summary
## --------------------------------------------------------------
## R                       0.938       RMSE              2.698
## R-Squared               0.879       Coef. Var         3.790
## Adj. R-Squared          0.877       MSE               7.277
## Pred R-Squared          0.870       MAE               2.062
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
```

```
##  MAE: Mean Absolute Error
##
##                            ANOVA
## -----------------------------------------------------------------
##              Sum of
##              Squares       DF     Mean Square       F        Sig.
## -----------------------------------------------------------------
## Regression  8529.407        3        2843.136   390.712    0.0000
## Residual    1171.566      161           7.277
## Total       9700.973      164
## -----------------------------------------------------------------
##
##                        Parameter Estimates
## ----------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t       Sig    lower     upper
## ----------------------------------------------------------------------------
## (Intercept)    47.908       1.649                   29.054   0.000   44.651    51.164
##          U     34.393       1.897        0.668      18.131   0.000   30.647    38.139
##          E     -0.024       0.003       -0.295      -8.255   0.000   -0.029    -0.018
##          I      0.042       0.009        0.134       4.630   0.000    0.024     0.061
## ----------------------------------------------------------------------------
##
##
##
## Stepwise Selection: Step 4
##
## - P added
##
##                        Model Summary
## ----------------------------------------------------------
## R                   0.942      RMSE             2.609
## R-Squared           0.888      Coef. Var        3.666
## Adj. R-Squared      0.885      MSE              6.809
## Pred R-Squared      0.876      MAE              1.955
## ----------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## -----------------------------------------------------------------
##              Sum of
##              Squares       DF     Mean Square       F        Sig.
## -----------------------------------------------------------------
## Regression  8611.466        4        2152.867   316.16     0.0000
## Residual    1089.507      160           6.809
## Total       9700.973      164
## -----------------------------------------------------------------
##
##                        Parameter Estimates
## ----------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t       Sig    lower     upper
## ----------------------------------------------------------------------------
## (Intercept)    48.601       1.608                   30.234   0.000   45.426    51.776
```

```
##              U     34.053           1.838         0.661     18.532   0.000   30.424    37.682
##              E     -0.018           0.003        -0.228     -5.790   0.000   -0.024    -0.012
##              I      0.033           0.009         0.103      3.499   0.001    0.014     0.051
##              P     -0.683           0.197        -0.124     -3.471   0.001   -1.071    -0.294
## --------------------------------------------------------------------------------------------
##
##
##
##                              Model Summary
## -----------------------------------------------------------------
## R                       0.942        RMSE                  2.609
## R-Squared               0.888        Coef. Var             3.666
## Adj. R-Squared          0.885        MSE                   6.809
## Pred R-Squared          0.876        MAE                   1.955
## -----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## -------------------------------------------------------------------------
##                  Sum of
##                  Squares        DF     Mean Square       F         Sig.
## -------------------------------------------------------------------------
## Regression      8611.466         4        2152.867    316.16      0.0000
## Residual        1089.507       160           6.809
## Total           9700.973       164
## -------------------------------------------------------------------------
##
##
##                           Parameter Estimates
## -----------------------------------------------------------------------------------------------
##        model      Beta    Std. Error    Std. Beta      t        Sig      lower      upper
## -----------------------------------------------------------------------------------------------
## (Intercept)     48.601         1.608                  30.234   0.000   45.426    51.776
##           U     34.053         1.838         0.661    18.532   0.000   30.424    37.682
##           E     -0.018         0.003        -0.228    -5.790   0.000   -0.024    -0.012
##           I      0.033         0.009         0.103     3.499   0.001    0.014     0.051
##           P     -0.683         0.197        -0.124    -3.471   0.001   -1.071    -0.294
## -----------------------------------------------------------------------------------------------
##
##
##
## No more variables to be added/removed.
##
##
## Final Model Output
## ------------------
##
##                              Model Summary
## -----------------------------------------------------------------
## R                       0.942        RMSE                  2.609
## R-Squared               0.888        Coef. Var             3.666
## Adj. R-Squared          0.885        MSE                   6.809
## Pred R-Squared          0.876        MAE                   1.955
```

```
## --------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                            ANOVA
## ----------------------------------------------------------------------
##             Sum of
##             Squares        DF     Mean Square      F        Sig.
## ----------------------------------------------------------------------
## Regression   8611.466       4       2152.867     316.16    0.0000
## Residual     1089.507     160          6.809
## Total        9700.973     164
## ----------------------------------------------------------------------
##
##
##                       Parameter Estimates
## ------------------------------------------------------------------------------
##       model      Beta    Std. Error   Std. Beta      t        Sig     lower      upper
## ------------------------------------------------------------------------------
## (Intercept)    48.601       1.608                  30.234    0.000    45.426     51.776
##         U      34.053       1.838      0.661       18.532    0.000    30.424     37.682
##         E      -0.018       0.003     -0.228       -5.790    0.000    -0.024     -0.012
##         I       0.033       0.009      0.103        3.499    0.001     0.014      0.051
##         P      -0.683       0.197     -0.124       -3.471    0.001    -1.071     -0.294
## ------------------------------------------------------------------------------
```

`summary(who.bothmodel$model) #adj.r.squared 0.8849  #E,I,P,U left`

```
##
## Call:
## lm(formula = paste(response, "~", paste(preds, collapse = " + ")),
##     data = l)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9589 -1.5002 -0.0549  1.7899  8.6629
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.600964   1.607515  30.234  < 2e-16 ***
## U           34.053141   1.837554  18.532  < 2e-16 ***
## E           -0.018245   0.003151  -5.790 3.61e-08 ***
## I            0.032575   0.009309   3.499 0.000604 ***
## P           -0.682684   0.196658  -3.471 0.000666 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.609 on 160 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8877, Adjusted R-squared:  0.8849
## F-statistic: 316.2 on 4 and 160 DF,  p-value: < 2.2e-16
```

```
#only backwardmodel has 6 predictors,analyze through assumption
who.backwardmodel2 <- lm(D~E+F+I+L+P+U,data=who.data.2015)
testAssumption(who.backwardmodel2) #not pass assumption test
```

```
## Loading required package: zoo

##
## Attaching package: 'zoo'

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

## Equal Variance pvalue:  0.2239423
## Normality pvalue:  0.0427232
## VIF :

##          VIF detection
## E  2.228195        0
## F 90.069777        1
## I  1.241236        0
## L 91.027456        1
## P  1.858764        0
## U  1.852147        0
```
```r
#delete F
who.backwardmodel3 <- lm(D~E+I+L+P+U,data=who.data.2015)
testAssumption(who.backwardmodel3) #  pass assumption test
```
```
## Equal Variance pvalue:  0.09808863
## Normality pvalue:  0.06610002
## VIF :

##          VIF detection
## E 2.226454        0
## I 1.237345        0
## L 1.057408        0
## P 1.832546        0
## U 1.844009        0
```
```r
#delete L
who.backwardmodel4 <- lm(D~E+F+I+P+U,data=who.data.2015)
testAssumption(who.backwardmodel4) #  pass assumption test
```
```
## Equal Variance pvalue:  0.0852452
## Normality pvalue:  0.07204847
## VIF :

##          VIF detection
## E 2.225540        0
## F 1.046283        0
## I 1.236752        0
## P 1.834507        0
## U 1.840313        0
```
```r
summary(who.backwardmodel3)  #0.8851 ,but p value for L is 0.260898 higher 0.05
```
```
##
## Call:
## lm(formula = D ~ E + I + L + P + U, data = who.data.2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -8.8958 -1.5039 -0.0622  1.7545  8.6227
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.859986   1.622469  30.115  < 2e-16 ***
## E           -0.018023   0.003154  -5.713 5.31e-08 ***
## I            0.032142   0.009309   3.453 0.000711 ***
## L           -0.002132   0.001889  -1.128 0.260898
## P           -0.690640   0.196617  -3.513 0.000578 ***
## U           33.780499   1.851821  18.242  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.607 on 159 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8886, Adjusted R-squared:  0.8851
## F-statistic: 253.6 on 5 and 159 DF,  p-value: < 2.2e-16
```

`summary`(who.backwardmodel4) *#0.8847 ,but p value for F is 0.369035 higher 0.05*

```
##
## Call:
## lm(formula = D ~ E + F + I + P + U, data = who.data.2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9224 -1.5281 -0.0647  1.7523  8.6422
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.795943   1.622960  30.066  < 2e-16 ***
## E           -0.018077   0.003158  -5.724 5.06e-08 ***
## F           -0.002181   0.002421  -0.901 0.369035
## I            0.032282   0.009320   3.464 0.000685 ***
## P           -0.691295   0.197006  -3.509 0.000585 ***
## U           33.848377   1.852634  18.270  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.611 on 159 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8883, Adjusted R-squared:  0.8847
## F-statistic: 252.8 on 5 and 159 DF,  p-value: < 2.2e-16
```

#also use subset method, to verify that the best predictors are 4 varaibles E,I,P,U
`library`(leaps)

```
## Warning: package 'leaps' was built under R version 4.0.4
```

best.subset<-`regsubsets`(D~`factor`(C)+E+F+I+J+K+L+M+O+P+Q+R+S+T+U+V, data= who.data.2015, nv=10)
`summary`(best.subset)

```
## Subset selection object
## Call: regsubsets.formula(D ~ factor(C) + E + F + I + J + K + L + M +
##     O + P + Q + R + S + T + U + V, data = who.data.2015, nv = 10)
## 16 Variables  (and intercept)
```

```
##                      Forced in Forced out
## factor(C)Developing     FALSE      FALSE
## E                       FALSE      FALSE
## F                       FALSE      FALSE
## I                       FALSE      FALSE
## J                       FALSE      FALSE
## K                       FALSE      FALSE
## L                       FALSE      FALSE
## M                       FALSE      FALSE
## O                       FALSE      FALSE
## P                       FALSE      FALSE
## Q                       FALSE      FALSE
## R                       FALSE      FALSE
## S                       FALSE      FALSE
## T                       FALSE      FALSE
## U                       FALSE      FALSE
## V                       FALSE      FALSE
## 1 subsets of each size up to 10
## Selection Algorithm: exhaustive
##           factor(C)Developing E   F   I   J   K   L   M   O   P   Q   R   S
## 1  ( 1 )  " "                 " " " " " " " " " " " " " " " " " " " " " " " "
## 2  ( 1 )  " "                 "*" " " " " " " " " " " " " " " " " " " " " " "
## 3  ( 1 )  " "                 "*" " " "*" " " " " " " " " " " " " " " " " " "
## 4  ( 1 )  " "                 "*" " " "*" " " " " " " " " " " "*" " " " " " "
## 5  ( 1 )  " "                 "*" " " "*" " " " " " " " " " " "*" " " " " "*"
## 6  ( 1 )  " "                 "*" "*" "*" " " " " " " "*" " " " " "*" " " " " "
## 7  ( 1 )  " "                 "*" "*" "*" " " " " "*" " " " " " " "*" " " " " "*"
## 8  ( 1 )  " "                 "*" "*" "*" " " " " "*" "*" " " " " "*" " " " " "*"
## 9  ( 1 )  " "                 "*" "*" "*" "*" " " "*" "*" " " " " "*" " " " " "*"
## 10 ( 1 )  " "                 "*" "*" "*" "*" "*" "*" "*" " " " " "*" " " " " "*"
##         T   U   V
## 1  ( 1 ) " " "*" " "
## 2  ( 1 ) " " "*" " "
## 3  ( 1 ) " " "*" " "
## 4  ( 1 ) " " "*" " "
## 5  ( 1 ) " " "*" " "
## 6  ( 1 ) " " "*" " "
## 7  ( 1 ) " " "*" " "
## 8  ( 1 ) " " "*" " "
## 9  ( 1 ) " " "*" " "
## 10 ( 1 ) " " "*" " "
```

```r
reg.summary <-summary(best.subset)
rsquare      <-c(reg.summary$rsq)
cp           <-c(reg.summary$cp)
AdjustedR    <-c(reg.summary$adjr2)
RMSE         <-c(reg.summary$rss)
cbind(rsquare,cp,RMSE,AdjustedR)
```

```
##        rsquare        cp       RMSE AdjustedR
## [1,] 0.8065098 93.836791 1596.4096 0.8049981
## [2,] 0.8696922 24.051144 1075.1170 0.8676401
## [3,] 0.8881239  5.109743  923.0448 0.8854601
## [4,] 0.8932890  1.241262  880.4291 0.8898743
## [5,] 0.8947742  1.553878  868.1757 0.8905312
```

```
## [6,] 0.8965342  1.554191   853.6544 0.8914871
## [7,] 0.8983486  1.492754   838.6847 0.8925161
## [8,] 0.8992035  2.521490   831.6315 0.8925392
## [9,] 0.8999510  3.672170   825.4639 0.8924473
## [10,] 0.9001967  5.393042   823.4370 0.8918098
```

```
#the first order model
who.firstordermodel <- lm(D~E+I+P+U, data = who.data.2015)
summary(who.firstordermodel) #0.8849
```

```
##
## Call:
## lm(formula = D ~ E + I + P + U, data = who.data.2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.9589 -1.5002 -0.0549  1.7899  8.6629
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 48.600964   1.607515  30.234  < 2e-16 ***
## E           -0.018245   0.003151  -5.790 3.61e-08 ***
## I            0.032575   0.009309   3.499 0.000604 ***
## P           -0.682684   0.196658  -3.471 0.000666 ***
## U           34.053141   1.837554  18.532  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.609 on 160 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8877, Adjusted R-squared:  0.8849
## F-statistic: 316.2 on 4 and 160 DF,  p-value: < 2.2e-16
```

```
testAssumption(who.firstordermodel)
```

```
## Equal Variance pvalue:  0.03293662
## Normality pvalue:  0.08693769
## VIF :
```

```
##       VIF detection
## E 2.217771         0
## I 1.235236         0
## P 1.830189         0
## U 1.812610         0
```

```
#boxcox
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.0.4
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:olsrr':
##
##     cement
```

```
bc=boxcox(who.firstordermodel,lambda=seq(-5,5))
```
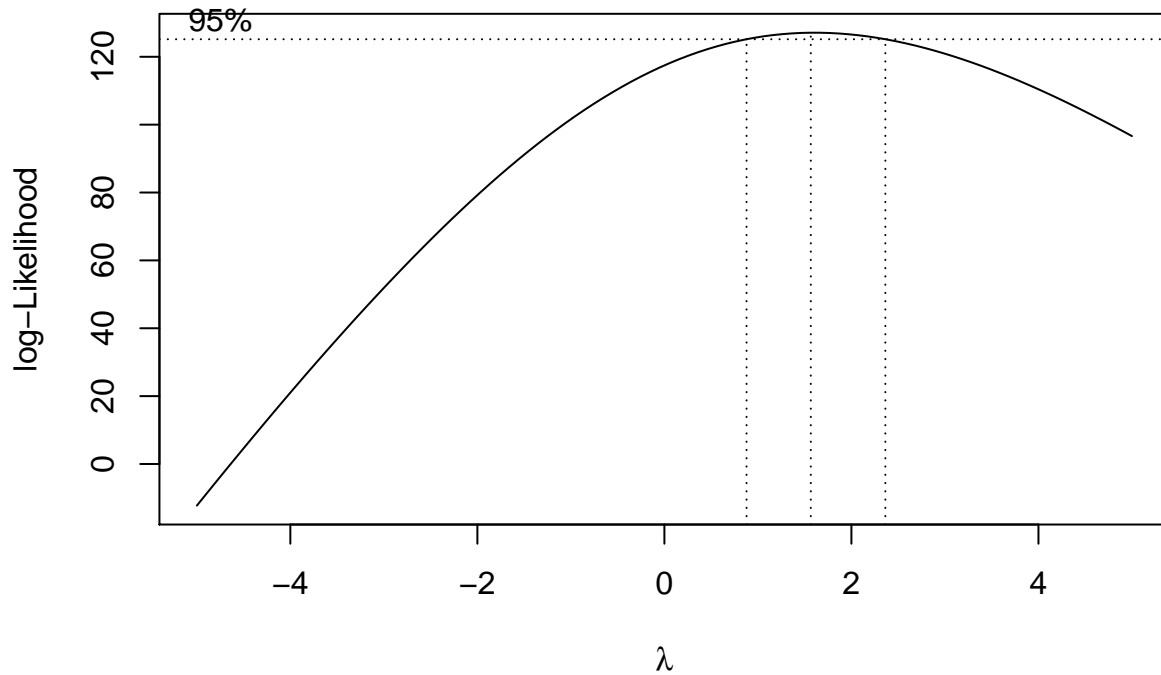


```
bestlambda=bc$x[which(bc$y==max(bc$y))]
bestlambda
```

```
## [1] 1.565657
```

```
who.firstordermodel.boxcox<- lm(((D^1.565657)-1)/1.565657~E+I+P+U  ,data = who.data.2015 )
summary(who.firstordermodel.boxcox) #0.8837
```

```
##
## Call:
## lm(formula = ((D^1.565657) - 1)/1.565657 ~ E + I + P + U, data = who.data.2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -88.649 -16.947  -0.027  19.833  88.211
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 256.00288   17.74019  14.431  < 2e-16 ***
## E            -0.19794    0.03477  -5.692 5.84e-08 ***
## I             0.35909    0.10273   3.496 0.000612 ***
## P            -6.58795    2.17027  -3.036 0.002804 **
## U           380.48616   20.27885  18.763  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## 
## Residual standard error: 28.8 on 160 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8865, Adjusted R-squared:  0.8837
## F-statistic: 312.5 on 4 and 160 DF,  p-value: < 2.2e-16
```

```
#test assumption again
testAssumption(who.firstordermodel.boxcox )
```

```
## Equal Variance pvalue:  0.08840221
## Normality pvalue:  0.3130584
## VIF :

##           VIF detection
## E 2.217771          0
## I 1.235236          0
## P 1.830189          0
## U 1.812610          0
```

```
plot(who.firstordermodel.boxcox,which=5)
```


Residuals vs Leverage

lm(((D^1.565657) − 1)/1.565657 ~ E + I + P + U)
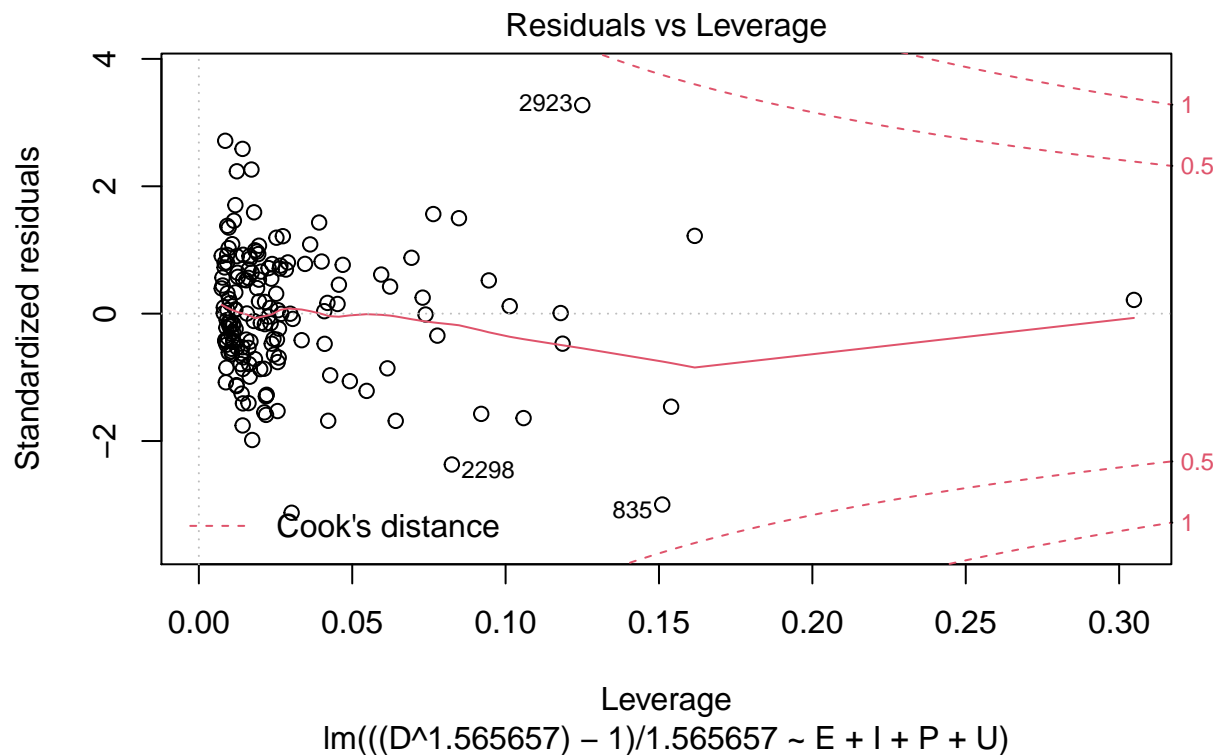
The first order model is who.firstordermodel.boxcox

```
who.firstordermodel.boxcox
```

```
## 
## Call:
## lm(formula = ((D^1.565657) - 1)/1.565657 ~ E + I + P + U, data = who.data.2015)
```

```
##
## Coefficients:
## (Intercept)            E            I            P            U
##     256.0029      -0.1979       0.3591      -6.5879     380.4862
```

```
summary(who.firstordermodel.boxcox)$adj.r.squared
```

```
## [1] 0.8836898
```

step 3 test to add interaction terms into model

```
who.interactionmodel <- lm(D~(E+I+P+U)^2,data = who.data.2015 )
summary(who.interactionmodel) #adj.r.squared 0.8939     E:I
```

```
##
## Call:
## lm(formula = D ~ (E + I + P + U)^2, data = who.data.2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.7152  -1.5042  -0.0437   1.5865   8.6678
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept) 38.8946788  5.4621161    7.121 3.84e-11 ***
## E           -0.0034061  0.0130164   -0.262  0.79392
## I            0.1724124  0.0632683    2.725  0.00717 **
## P            0.1408767  1.7092613    0.082  0.93442
## U           46.4706998  7.4400038    6.246 3.92e-09 ***
## E:I         -0.0002520  0.0001092   -2.307  0.02238 *
## E:P          0.0018817  0.0016690    1.127  0.26130
## E:U         -0.0012967  0.0196214   -0.066  0.94740
## I:P          0.0017069  0.0067289    0.254  0.80009
## I:U         -0.1623697  0.0825328   -1.967  0.05094 .
## P:U         -2.8453478  2.6225967   -1.085  0.27965
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.505 on 154 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.9004, Adjusted R-squared:  0.8939
## F-statistic: 139.2 on 10 and 154 DF,  p-value: < 2.2e-16
```

```
who.interactionmodel2 <- lm(D~E+I+P+U+E*I  ,data = who.data.2015 )
summary(who.interactionmodel2) #adj.r.squared  0.8891     better than first order model 0.8837
```

```
##
## Call:
## lm(formula = D ~ E + I + P + U + E * I, data = who.data.2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.1017  -1.6166  -0.0114   1.7045   9.2466
##
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.677e+01  1.724e+00  27.131  < 2e-16 ***
## E           -4.651e-03  5.994e-03  -0.776 0.438980
## I            6.666e-02  1.579e-02   4.222 4.06e-05 ***
## P           -6.833e-01  1.931e-01  -3.539 0.000527 ***
## U            3.310e+01  1.840e+00  17.992  < 2e-16 ***
## E:I         -1.982e-04  7.486e-05  -2.648 0.008920 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.562 on 159 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8924, Adjusted R-squared:  0.8891
## F-statistic: 263.8 on 5 and 159 DF,  p-value: < 2.2e-16
```

```
#test assumption
testAssumption(who.interactionmodel2)   # not pass Normality pvalue:  0.0461
```

```
## Equal Variance pvalue:  0.06104085
## Normality pvalue:  0.04608459
## VIF :

##          VIF detection
## E   8.327457          0
## I   3.686859          0
## P   1.830191          0
## U   1.884996          0
## E:I 8.476546          0
```

```
#box-cox ,transformation
bc=boxcox(who.interactionmodel2,lambda=seq(-5,5))
```
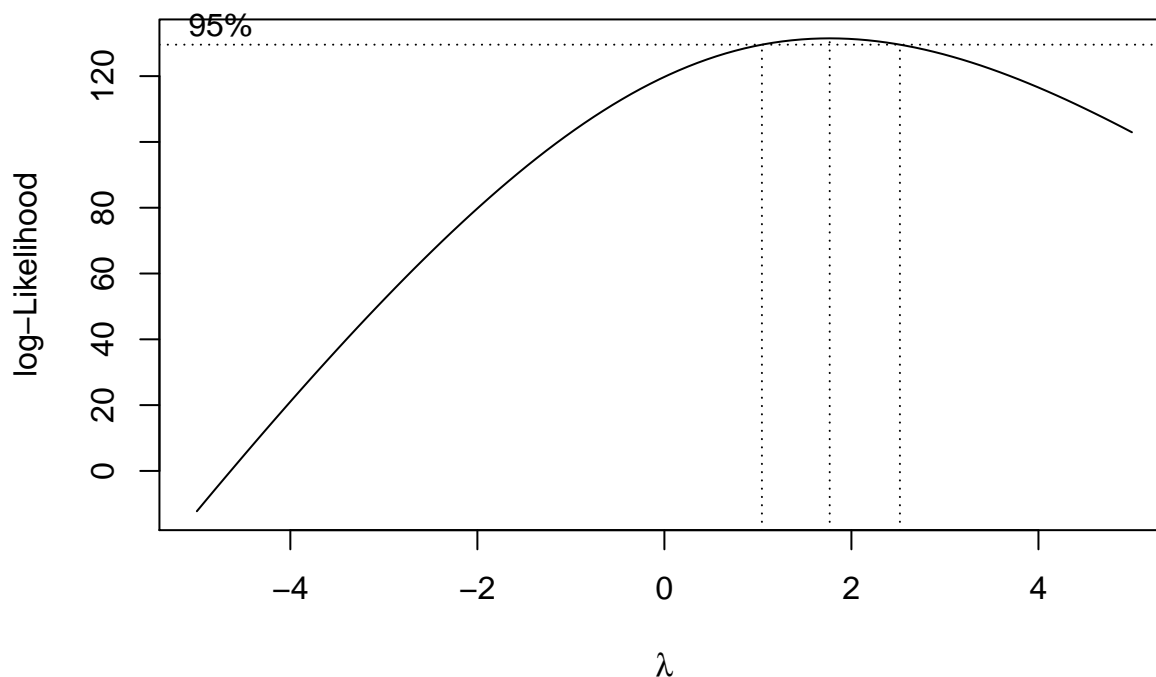
```r
bestlambda=bc$x[which(bc$y==max(bc$y))]
bestlambda
```

```
## [1] 1.767677
```

```r
who.interactionmodel3<- lm(((D^1.767677)-1)/1.767677~E+I+P+U+E*I  ,data = who.data.2015 )
summary(who.interactionmodel3) #0.8882
```

```
##
## Call:
## lm(formula = ((D^1.767677) - 1)/1.767677 ~ E + I + P + U + E *
##     I, data = who.data.2015)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -205.13  -43.37    0.50   42.05  219.07
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 416.785962  44.743847   9.315  < 2e-16 ***
## E            -0.067370   0.155599  -0.433  0.66562
## I             1.842222   0.409835   4.495 1.33e-05 ***
## P           -14.768409   5.011665  -2.947  0.00369 **
## U           873.469276  47.754408  18.291  < 2e-16 ***
## E:I          -0.005789   0.001943  -2.979  0.00334 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 66.5 on 159 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8916, Adjusted R-squared:  0.8882
## F-statistic: 261.6 on 5 and 159 DF,  p-value: < 2.2e-16
```

```
#test assumption again
testAssumption(who.interactionmodel3) #pass assumption test
```

```
## Equal Variance pvalue:  0.1720505
## Normality pvalue:  0.2053997
## VIF :
```

```
##            VIF detection
## E   8.327457          0
## I   3.686859          0
## P   1.830191          0
## U   1.884996          0
## E:I 8.476546          0
```

```
plot(who.interactionmodel3,which=5) # no outlier
```


Residuals vs Leverage

lm(((D^1.767677) − 1)/1.767677 ~ E + I + P + U + E * I)
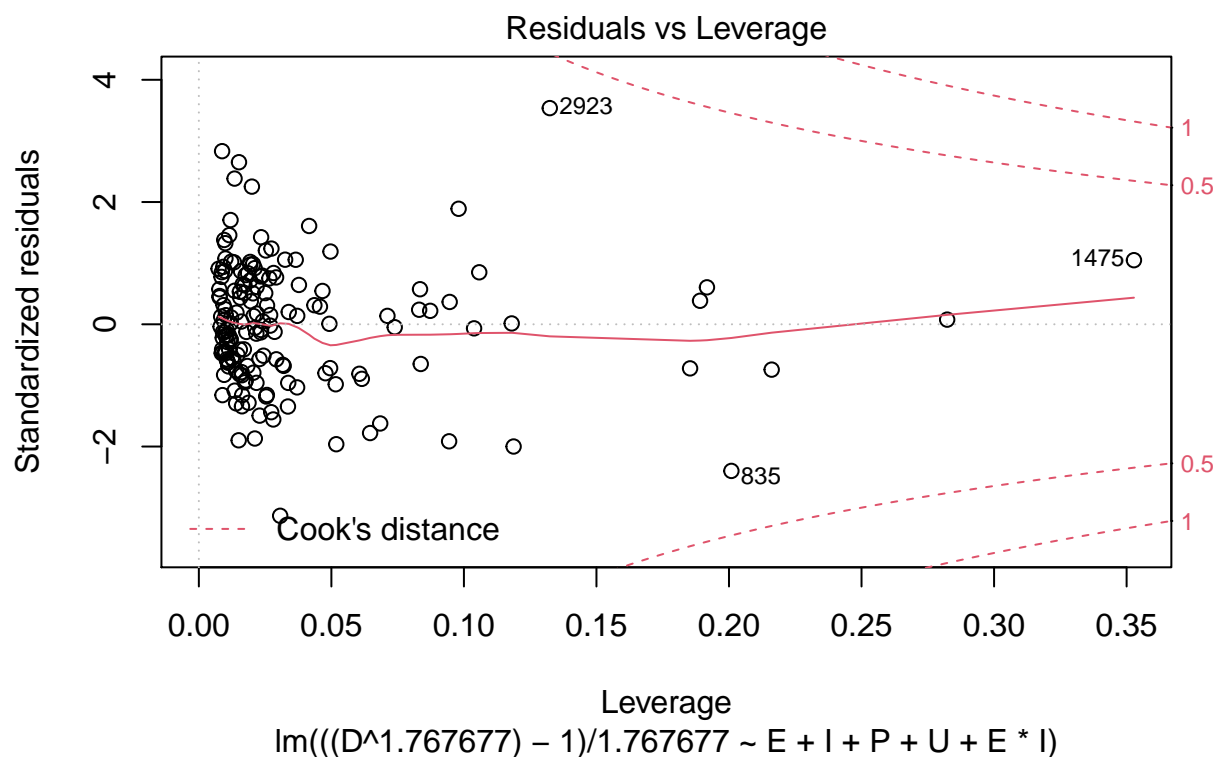
The interaction terms model is who.interactionmodel3

```
who.interactionmodel3
```

```
##
## Call:
```

```
## lm(formula = ((D^1.767677) - 1)/1.767677 ~ E + I + P + U + E *
##     I, data = who.data.2015)
##
## Coefficients:
## (Intercept)            E            I            P            U          E:I
##  416.785962    -0.067370     1.842222   -14.768409   873.469276    -0.005789
```

```r
summary(who.interactionmodel3)$adj.r.squared
```

```
## [1] 0.8882241
```

step 4 test to add high order

```r
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 4.0.4
```

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

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

```r
#who.data.2015.sub <-who.data.2015[c("D","E","I","P","U")]
pairs(~D +E +I +P +U,data=who.data.2015 )
```
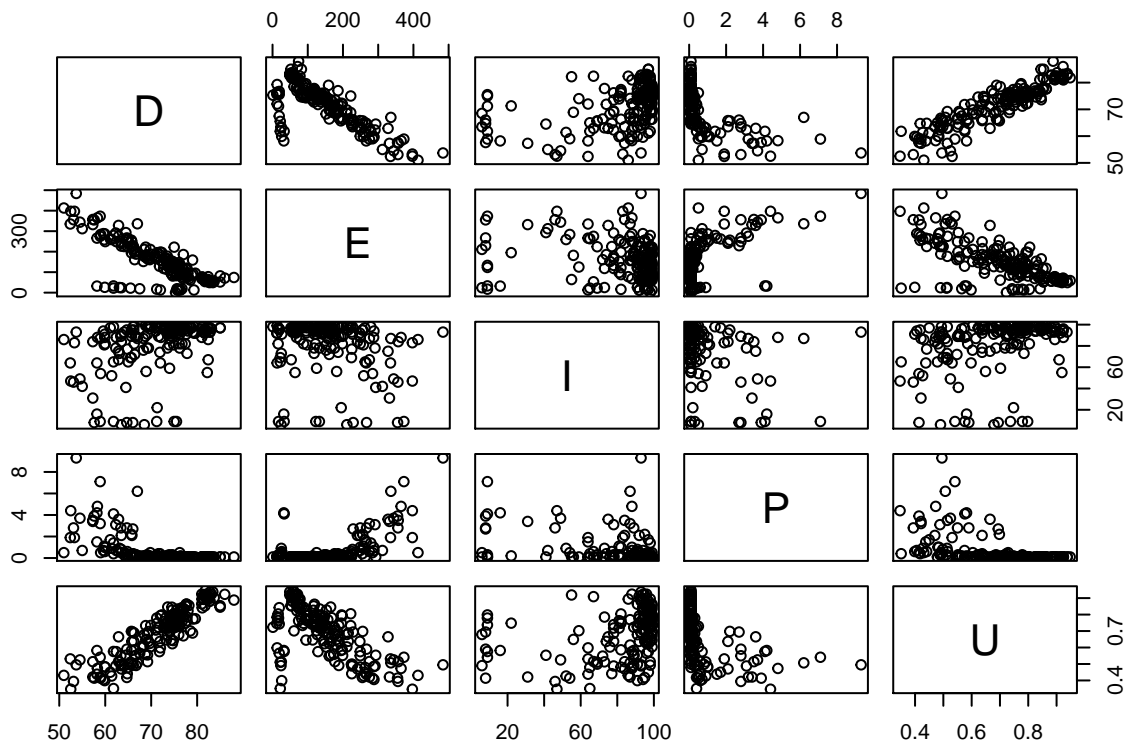


```r
who.quadmodelU  <- lm(D~E +I +P +U +I(U^2) +E*I      ,data = who.data.2015)
summary(who.quadmodelU)  #adj.r.squared 0.8884   , dcreased little
```

```
##
## Call:
## lm(formula = D ~ E + I + P + U + I(U^2) + E * I, data = who.data.2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.1067  -1.6267  -0.0258   1.7132   9.2422
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.646e+01  4.321e+00   10.753  < 2e-16 ***
## E           -4.589e-03  6.066e-03   -0.757 0.450412
## I            6.686e-02  1.604e-02    4.169 5.03e-05 ***
## P           -6.814e-01  1.953e-01   -3.489 0.000628 ***
## U            3.404e+01  1.236e+01    2.753 0.006596 **
## I(U^2)      -7.168e-01  9.336e+00   -0.077 0.938901
## E:I         -1.993e-04  7.648e-05   -2.606 0.010033 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.57 on 158 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8924, Adjusted R-squared:  0.8884
## F-statistic: 218.5 on 6 and 158 DF,  p-value: < 2.2e-16
```

```r
who.quadmodelE  <- lm(D~E +I +P +U +I(E^2) +E*I      ,data = who.data.2015)
summary(who.quadmodelE)  #adj.r.squared 0.896 , increased
```

```
##
## Call:
## lm(formula = D ~ E + I + P + U + I(E^2) + E * I, data = who.data.2015)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.0598  -1.4883  -0.0657   1.5514   7.9960
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.535e+01  1.720e+00   26.369  < 2e-16 ***
## E            2.022e-02  9.322e-03    2.169 0.031551 *
## I            7.397e-02  1.543e-02    4.792 3.78e-06 ***
## P           -1.300e-01  2.475e-01   -0.525 0.600213
## U            3.228e+01  1.797e+00   17.964  < 2e-16 ***
## I(E^2)      -7.447e-05  2.184e-05   -3.410 0.000826 ***
## E:I         -2.302e-04  7.308e-05   -3.150 0.001956 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.48 on 158 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8998, Adjusted R-squared:  0.896
## F-statistic: 236.5 on 6 and 158 DF,  p-value: < 2.2e-16
```

```r
who.cubicE  <- lm(D~E +I +P +U +I(E^2) +I(E^3) +E*I      ,data = who.data.2015)
summary(who.cubicE)    #adj.r.squared 0.9027 ,best
```

```
## 
## Call:
## lm(formula = D ~ E + I + P + U + I(E^2) + I(E^3) + E * I, data = who.data.2015)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1450 -1.5834 -0.1062  1.4450  9.4275
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.576e+01  1.667e+00  27.443  < 2e-16 ***
## E            7.199e-02  1.748e-02   4.118 6.15e-05 ***
## I            7.579e-02  1.494e-02   5.074 1.09e-06 ***
## P           -1.865e-01  2.399e-01  -0.777 0.438137
## U            2.934e+01  1.936e+00  15.156  < 2e-16 ***
## I(E^2)      -3.744e-04  8.929e-05  -4.193 4.59e-05 ***
## I(E^3)       4.690e-07  1.357e-07   3.457 0.000704 ***
## E:I         -2.809e-04  7.218e-05  -3.892 0.000147 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.399 on 157 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.9069, Adjusted R-squared:  0.9027
## F-statistic: 218.5 on 7 and 157 DF,  p-value: < 2.2e-16
```

```r
who.4orderE  <- lm(D~E +I +P +U +I(E^2) +I(E^3) +I(E^4) +E*I        ,data = who.data.2015)
summary(who.4orderE)    #adj.r.squared 0.9026
```

```
## 
## Call:
## lm(formula = D ~ E + I + P + U + I(E^2) + I(E^3) + I(E^4) + E *
##     I, data = who.data.2015)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.3776 -1.5932 -0.0318  1.4296  9.1303
## 
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.577e+01  1.669e+00  27.433  < 2e-16 ***
## E            9.298e-02  2.947e-02   3.155 0.001927 **
## I            7.260e-02  1.537e-02   4.722 5.17e-06 ***
## P           -1.993e-01  2.405e-01  -0.829 0.408643
## U            2.893e+01  1.992e+00  14.521  < 2e-16 ***
## I(E^2)      -6.034e-04  2.738e-04  -2.204 0.028981 *
## I(E^3)       1.261e-06  9.054e-07   1.393 0.165548
## I(E^4)      -8.692e-10  9.820e-10  -0.885 0.377457
## E:I         -2.579e-04  7.678e-05  -3.358 0.000985 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.4 on 156 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.9074, Adjusted R-squared:  0.9026
```

```
## F-statistic:   191 on 8 and 156 DF,  p-value: < 2.2e-16
#test assumption
testAssumption(who.cubicE)  #best adj.r.squared, but do not pass assumption test

## Equal Variance pvalue:  0.004606545
## Normality pvalue:  0.01063238
## VIF :

##              VIF detection
## E        80.776821        1
## I         3.763924        0
## P         3.224464        0
## U         2.380690        0
## I(E^2) 336.635257         1
## I(E^3) 119.488075         1
## E:I       8.989846        0
#try box-cox transformation
bc=boxcox(who.cubicE,lambda=seq(-5,5))
```
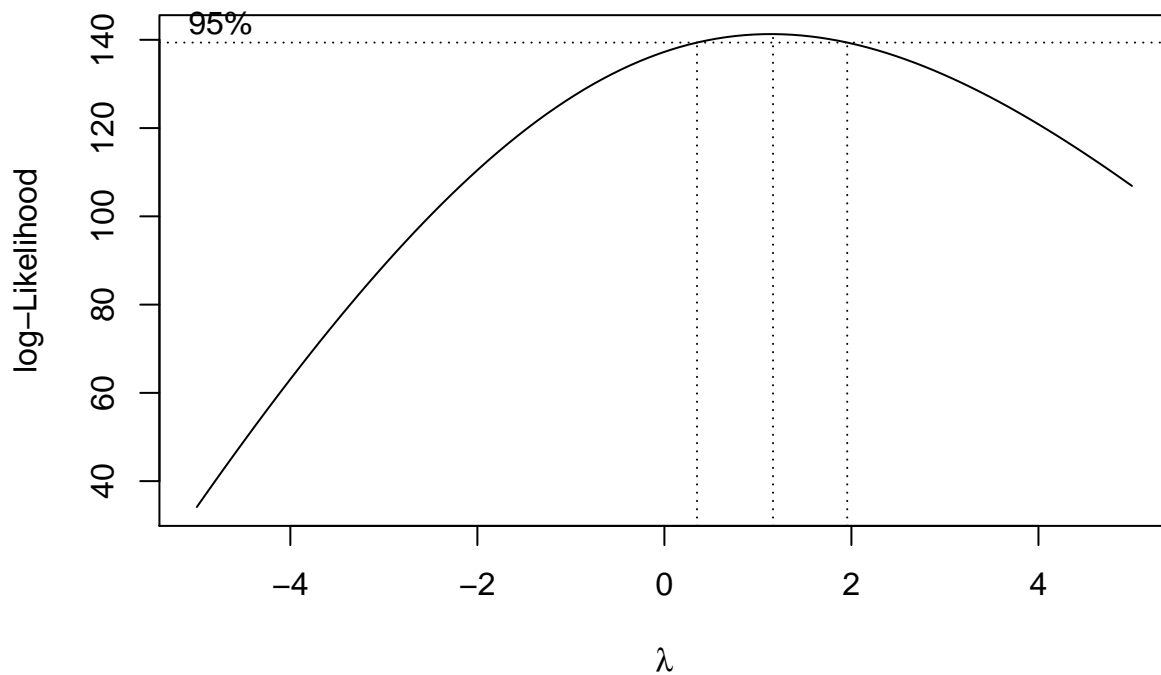


```
bestlambda=bc$x[which(bc$y==max(bc$y))]
bestlambda
```

```
## [1] 1.161616
```

```
who.cubicE2  <- lm(((D^1.161616)-1)/ 1.161616~E +I +P +U +I(E^2) +I(E^3) +E*I     ,data = who.data.2015)
summary(who.cubicE2)  #0.893
```

```
##
## Call:
## lm(formula = ((D^1.161616) - 1)/1.161616 ~ E + I + P + U + I(E^2) +
##     I(E^3) + E * I, data = who.data.2015)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -13.7885  -3.1806  -0.1761   2.8390  18.5390
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.076e+01  3.318e+00  21.329  < 2e-16 ***
## E            1.407e-01  3.478e-02   4.046 8.15e-05 ***
## I            1.512e-01  2.972e-02   5.089 1.02e-06 ***
## P           -3.969e-01  4.774e-01  -0.831 0.407060
## U            5.853e+01  3.851e+00  15.197  < 2e-16 ***
## I(E^2)      -7.348e-04  1.777e-04  -4.136 5.75e-05 ***
## I(E^3)       9.338e-07  2.700e-07   3.459 0.000699 ***
## E:I         -5.639e-04  1.436e-04  -3.926 0.000129 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.772 on 157 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.9062, Adjusted R-squared:  0.902
## F-statistic: 216.6 on 7 and 157 DF,  p-value: < 2.2e-16
```

```r
#test assumption
testAssumption(who.cubicE2)  #best adj.r.squared, but do not pass assumption test
```

```
## Equal Variance pvalue:  0.0065634
## Normality pvalue:  0.0132537
## VIF :
```

```
##               VIF detection
## E         80.776821         1
## I          3.763924         0
## P          3.224464         0
## U          2.380690         0
## I(E^2) 336.635257         1
## I(E^3) 119.488075         1
## E:I        8.989846         0
```

```r
#high order can not pass assumption test ,even after transformation
```

step 5 final model , compare adj.R.squared interaction terms order(0.888) is better than first order model(0.883) , now we name the model as ModelR

```r
ModelR <- who.interactionmodel3
coef(ModelR)
```

```
##   (Intercept)             E             I             P             U
## 416.785962465   -0.067370406   1.842221630 -14.768409390 873.469275608
```

```
##              E:I
##   -0.005789386
```

```
summary(ModelR)
```

```
##
## Call:
## lm(formula = ((D^1.767677) - 1)/1.767677 ~ E + I + P + U + E *
##     I, data = who.data.2015)
##
## Residuals:
##     Min     1Q  Median     3Q     Max
## -205.13  -43.37    0.50   42.05  219.07
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 416.785962  44.743847   9.315  < 2e-16 ***
## E            -0.067370   0.155599  -0.433  0.66562
## I             1.842222   0.409835   4.495 1.33e-05 ***
## P           -14.768409   5.011665  -2.947  0.00369 **
## U           873.469276  47.754408  18.291  < 2e-16 ***
## E:I          -0.005789   0.001943  -2.979  0.00334 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 66.5 on 159 degrees of freedom
##   (18 observations deleted due to missingness)
## Multiple R-squared:  0.8916, Adjusted R-squared:  0.8882
## F-statistic: 261.6 on 5 and 159 DF,  p-value: < 2.2e-16
```

$$\frac{(\widehat{Lifeexpectancy})^{1.767677}-1}{1.767677} = 416.786 - 0.067 * AdultMortality + 1.842 * HepatitisB - 14.768 * HIV/AIDS + 873.469 * Incomecompositionofresources - 0.006 * AdultMortality * HepatitisB$$