사회적 요인이 기대수명에 미치는 영향

R을 이용한 다중 선형 회귀 분석

통계학과
심정은
양수형
이혜진

목차

01 전처리

1. NA 제거
2. KNN
3. Data set

02 데이터 분석

1. Assumption
2. Outlier & Influential Point
3. Log transformation
4. Variable selection
5. Multiple collinearity
6. Model selection

03 한계점

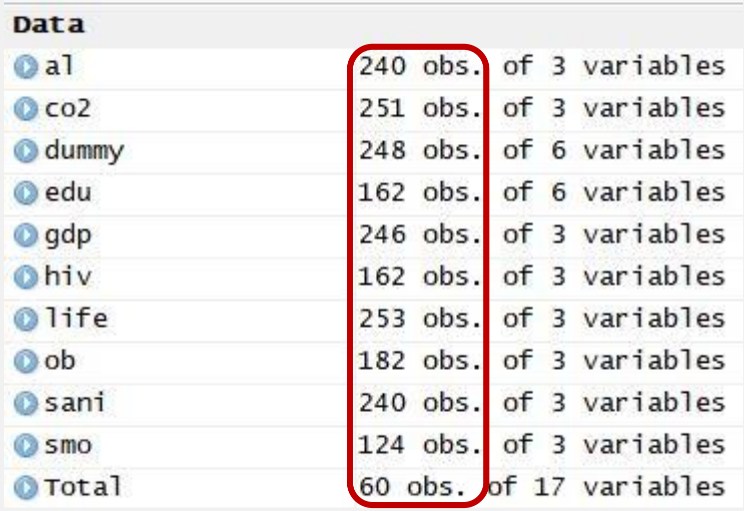
00 목적_

기대 수명과 사회적 요인 간의 관계 모델 적합.

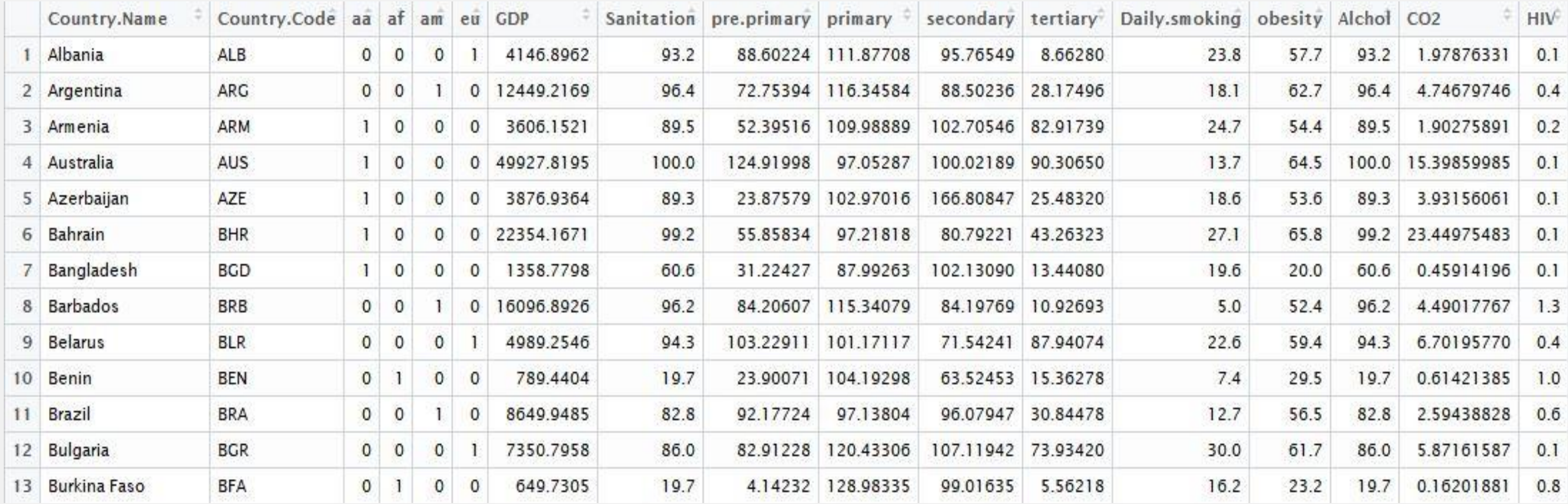
$$life_i = \beta_0 + \beta_1 gdp_i + \beta_2 sani_i + \beta_3 pre_i + \beta_4 pri_i + \beta_5 sec_i \\ + \beta_6 ter_i + \beta_7 smo_i + \beta_8 ob_i + \beta_9 al_i + \beta_{10} co2_i + \beta_{11} hiv_i$$

A. **life** : Life expectancy at birth, total(year) / 2015
B. **gdp** : GDP per capita (current US$) / 2016
C. **sani** : Improved sanitation facilities (% of population with access) / 2015
D. **pre** : Gross enrollment ratio, pre-primary, both sex (%) / 2015
E. **pri** : Gross enrollment ratio, primary, both sex (%) / 2015
F. **sec** : Gross enrollment ratio. secondary, both sex (%) / 2015
G. **ter** : Gross enrollment ratio. tertiary, both sex (%) / 2015
H. **smo** : smoking 2013 daily cigarette, both sexed, aged-standardized rate / 2013
I. **ob** : prevalence of obesity, BMI >= 25, 18+, age-standardized estimate / 2016
J. **al** : Total alcohol consumption per capita
(liters of pure alcohol, projected estimates, 15+ years of age) / 2015
K. **co2** : CO2 emissions (metric tons per capita) / 2014
L. **hiv** : Prevalence of HIV, total (% of population ages 15-49) / 2016

01 전처리_NA 제거



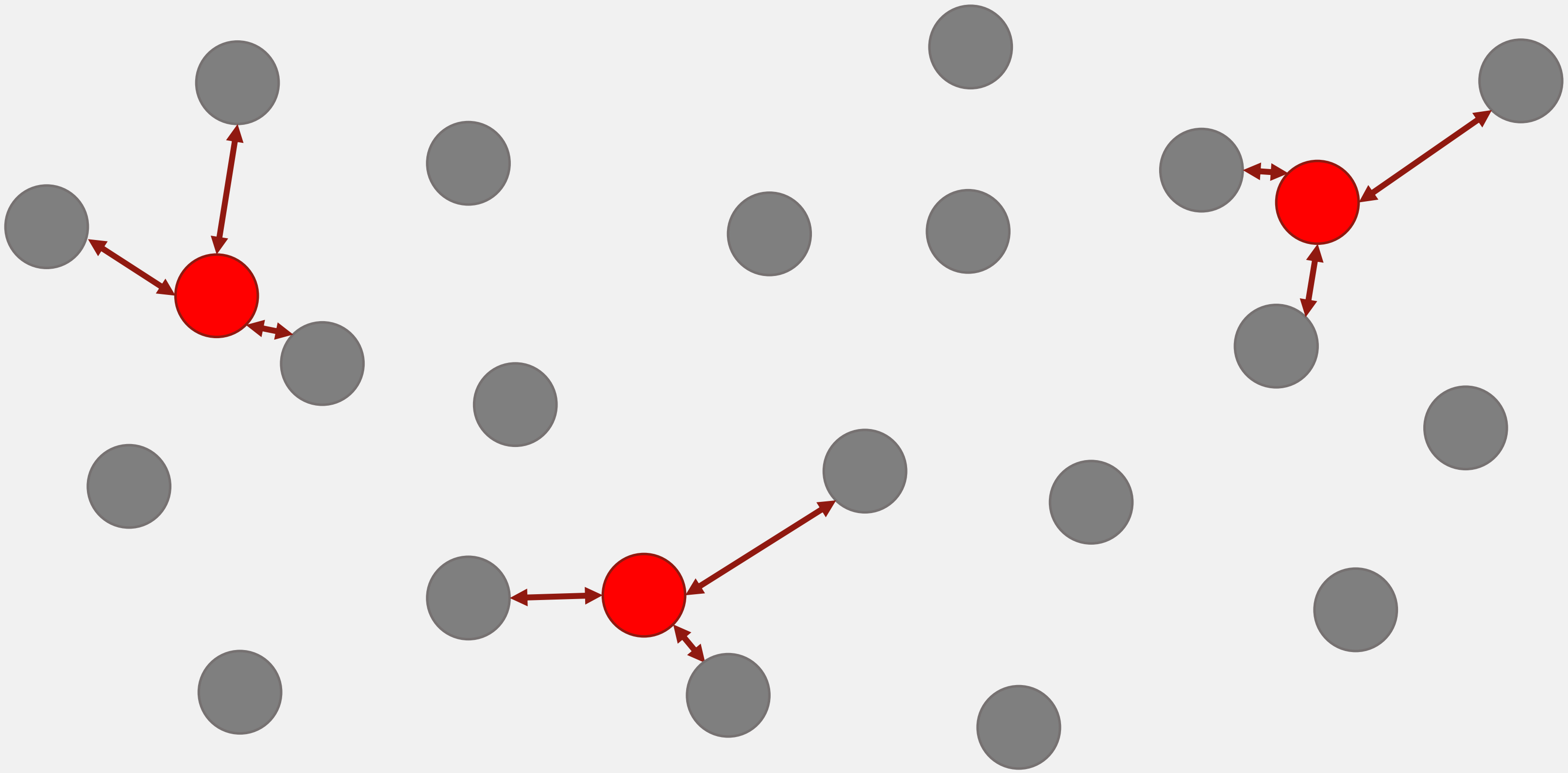
```
Data
al          240 obs. of 3 variables
co2         251 obs. of 3 variables
dummy       248 obs. of 6 variables
edu         162 obs. of 6 variables
gdp         246 obs. of 3 variables
hiv         162 obs. of 3 variables
life        253 obs. of 3 variables
ob          182 obs. of 3 variables
sani        240 obs. of 3 variables
smo         124 obs. of 3 variables
Total       60 obs. of 17 variables
```



	Country.Name	Country.Code	aa	af	am	eu	GDP	Sanitation	pre.primary	primary	secondary	tertiary	Daily.smoking	obesity	Alchol	CO2	HIV
1	Albania	ALB	0	0	0	1	4146.8962	93.2	88.60224	111.87708	95.76549	8.66280	23.8	57.7	93.2	1.97876331	0.1
2	Argentina	ARG	0	0	1	0	12449.2169	96.4	72.75394	116.34584	88.50236	28.17496	18.1	62.7	96.4	4.74679746	0.4
3	Armenia	ARM	1	0	0	0	3606.1521	89.5	52.39516	109.98889	102.70546	82.91739	24.7	54.4	89.5	1.90275891	0.2
4	Australia	AUS	1	0	0	0	49927.8195	100.0	124.91998	97.05287	100.02189	90.30650	13.7	64.5	100.0	15.39859985	0.1
5	Azerbaijan	AZE	1	0	0	0	3876.9364	89.3	23.87579	102.97016	166.80847	25.48320	18.6	53.6	89.3	3.93156061	0.1
6	Bahrain	BHR	1	0	0	0	22354.1671	99.2	55.85834	97.21818	80.79221	43.26323	27.1	65.8	99.2	23.44975483	0.1
7	Bangladesh	BGD	1	0	0	0	1358.7798	60.6	31.22427	87.99263	102.13090	13.44080	19.6	20.0	60.6	0.45914196	0.1
8	Barbados	BRB	0	0	1	0	16096.8926	96.2	84.20607	115.34079	84.19769	10.92693	5.0	52.4	96.2	4.49017767	1.3
9	Belarus	BLR	0	0	0	1	4989.2546	94.3	103.22911	101.17117	71.54241	87.94074	22.6	59.4	94.3	6.70195770	0.4
10	Benin	BEN	0	1	0	0	789.4404	19.7	23.90071	104.19298	63.52453	15.36278	7.4	29.5	19.7	0.61421385	1.0
11	Brazil	BRA	0	0	1	0	8649.9485	82.8	92.17724	97.13804	96.07947	30.84478	12.7	56.5	82.8	2.59438828	0.6
12	Bulgaria	BGR	0	0	0	1	7350.7958	86.0	82.91228	120.43306	107.11942	73.93420	30.0	61.7	86.0	5.87161587	0.1
13	Burkina Faso	BFA	0	1	0	0	649.7305	19.7	4.14232	128.98335	99.01635	5.56218	16.2	23.2	19.7	0.16201881	0.8

01 전처리_KNN (K-Nearest Neighbor)

3NN



01 전처리_KNN

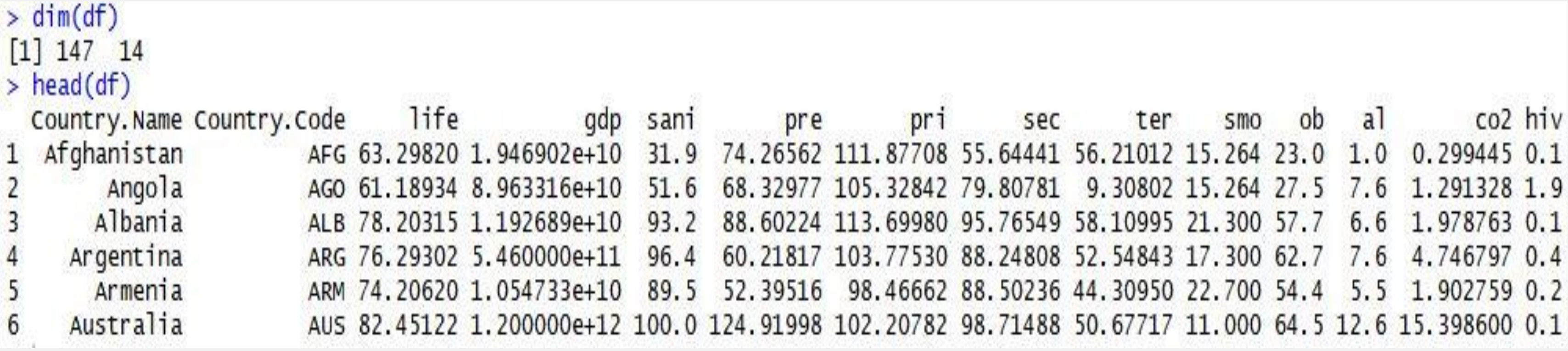
```
KNN <- function(data, year){
  if(length(which(is.na(data[,year]))) == 0) return(data[,c(1,2, year)])
  na.row <- which(is.na(data[, year]))

  for(i in 1:length(na.row)){
    col <- !is.na(data[na.row[i],])
    collected.col <- data[, col]; ncol <- length(collected.col)-2
    key <- collected.col[na.row[i],]
    index <- complete.cases(collected.col)
    non.na <- collected.col[index,]
    d <- apply(as.data.frame(non.na[,-c(1,2)]), 1, "-", key[,-c(1,2)])
    d2 <- unlist(d) ^ 2
    d2.matrix <- as.data.frame(matrix(d2, length(d2)/ncol, ncol, byrow = T))
    colnames(d2.matrix) <- colnames(non.na)[-c(1, 2)]
    p.length <- apply(c, 1, sum)
    o.p <- order(p.length)
    n.point <- o.p[which(!is.na(data[o.p, year]))][1:5]
    n.data <- data[n.point, year]
    data[na.row[i], year] <- mean(n.data, na.rm = T)
  }
  return(data[,c(1, 2, year)])
}
```

key observation에서 NA가 아닌 열 추출

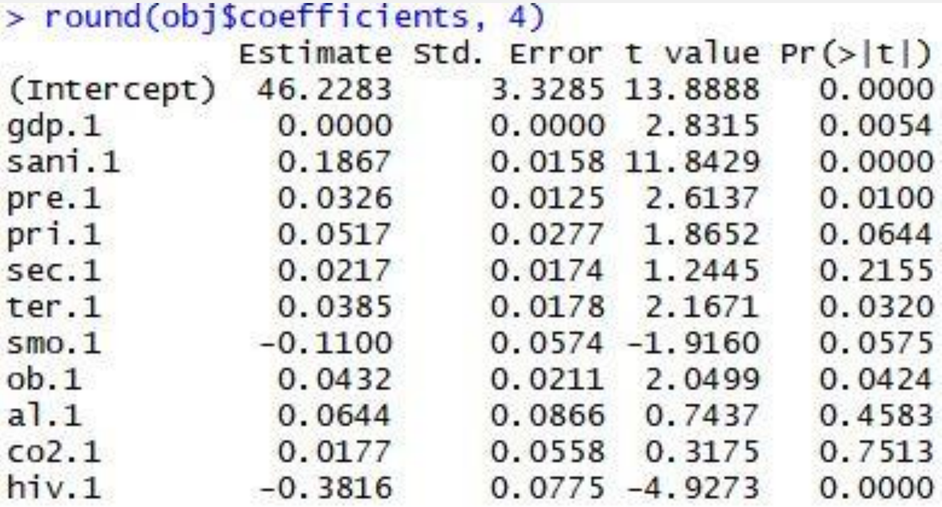
Euclidean distance

기준 년도에서 NA가
아닌 국가들 중,
key observation과 거리의
합이 가장 가까운
5개의 국가 선출.

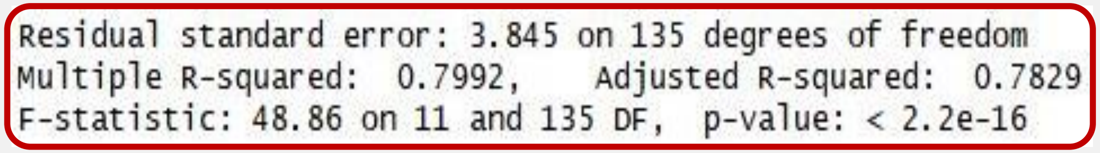
01 전처리_Data set



```
> dim(df)
[1] 147  14
> head(df)
  Country.Name Country.Code     life          gdp  sani       pre       pri      sec      ter    smo   ob   al       co2 hiv
1  Afghanistan          AFG 63.29820 1.946902e+10  31.9  74.26562 111.87708 55.64441 56.21012 15.264 23.0  1.0  0.299445 0.1
2       Angola          AGO 61.18934 8.963316e+10  51.6  68.32977 105.32842 79.80781  9.30802 15.264 27.5  7.6  1.291328 1.9
3      Albania          ALB 78.20315 1.192689e+10  93.2  88.60224 113.69980 95.76549 58.10995 21.300 57.7  6.6  1.978763 0.1
4    Argentina          ARG 76.29302 5.460000e+11  96.4  60.21817 103.77530 88.24808 52.54843 17.300 62.7  7.6  4.746797 0.4
5      Armenia          ARM 74.20620 1.054733e+10  89.5  52.39516  98.46662 88.50236 44.30950 22.700 54.4  5.5  1.902759 0.2
6    Australia          AUS 82.45122 1.200000e+12 100.0 124.91998 102.20782 98.71488 50.67717 11.000 64.5 12.6 15.398600 0.1
```



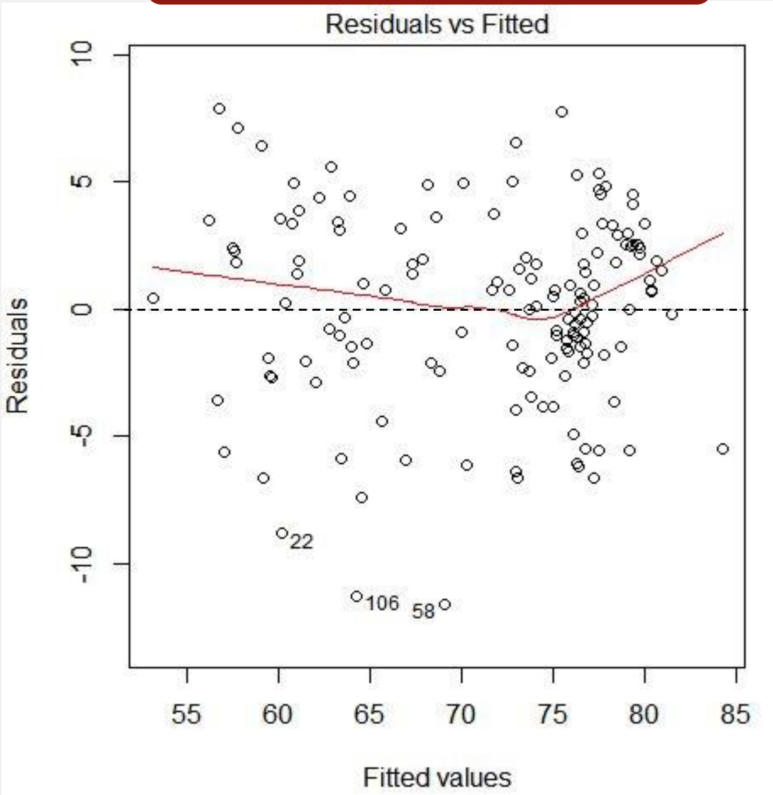
```
> round(obj$coefficients, 4)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  46.2283     3.3285 13.8888   0.0000
gdp.1         0.0000     0.0000  2.8315   0.0054
sani.1        0.1867     0.0158 11.8429   0.0000
pre.1         0.0326     0.0125  2.6137   0.0100
pri.1         0.0517     0.0277  1.8652   0.0644
sec.1         0.0217     0.0174  1.2445   0.2155
ter.1         0.0385     0.0178  2.1671   0.0320
smo.1        -0.1100     0.0574 -1.9160   0.0575
ob.1          0.0432     0.0211  2.0499   0.0424
al.1          0.0644     0.0866  0.7437   0.4583
co2.1         0.0177     0.0558  0.3175   0.7513
hiv.1        -0.3816     0.0775 -4.9273   0.0000
```



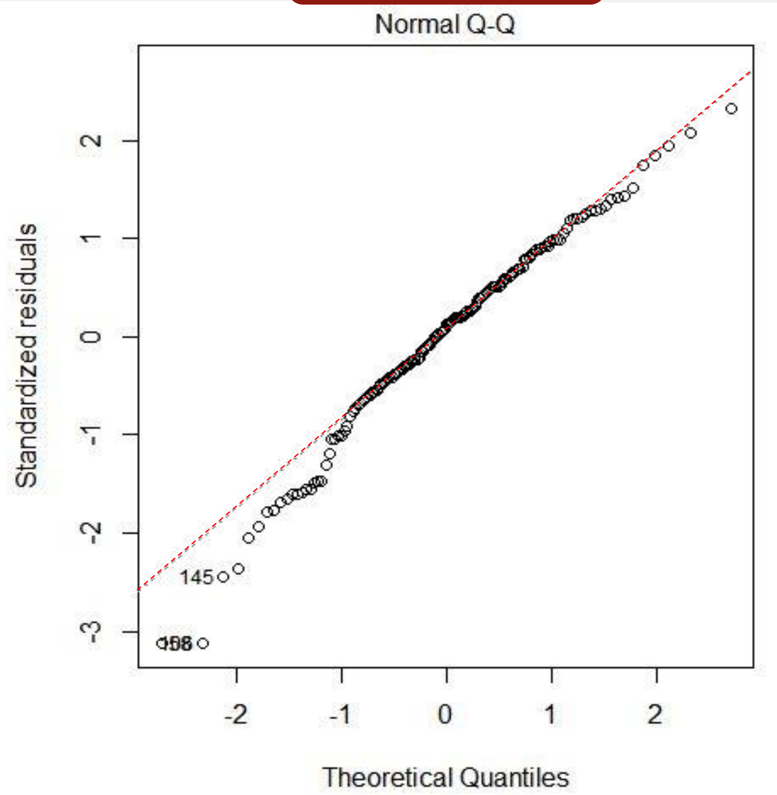
```
Residual standard error: 3.845 on 135 degrees of freedom
Multiple R-squared:  0.7992,    Adjusted R-squared:  0.7829
F-statistic: 48.86 on 11 and 135 DF,  p-value: < 2.2e-16
```

02 데이터 분석_Assumption test

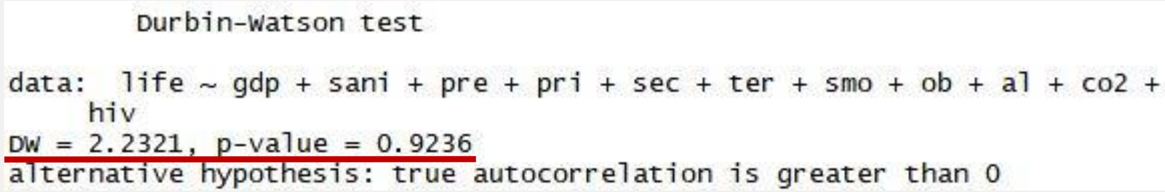
Heteroscedasticity



Normality



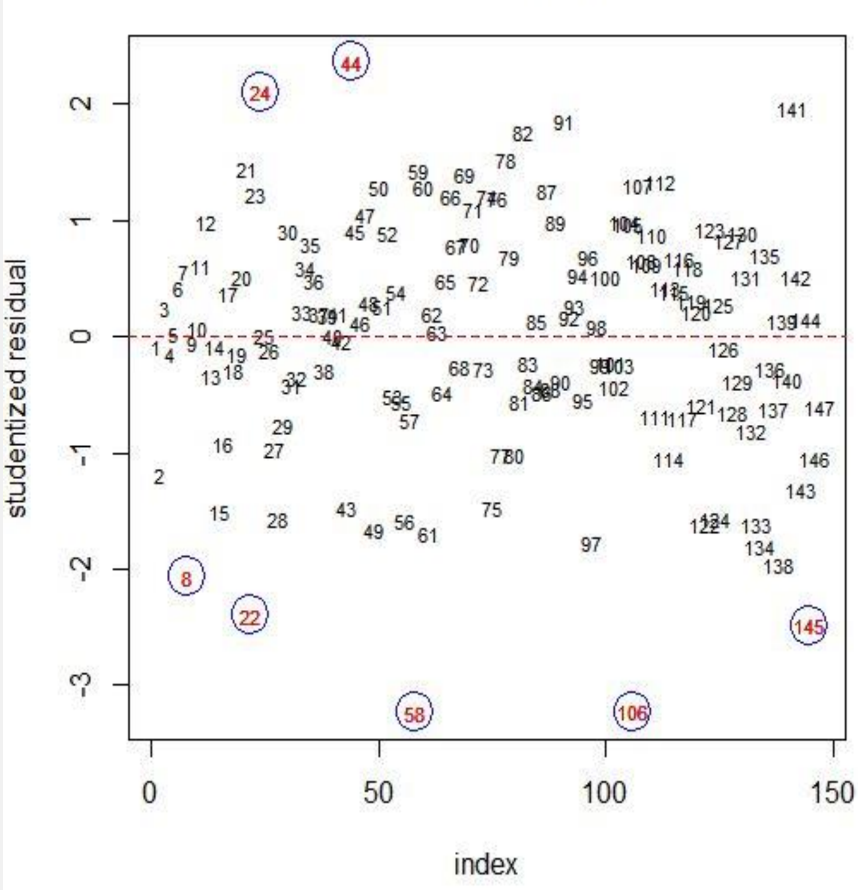
Independency



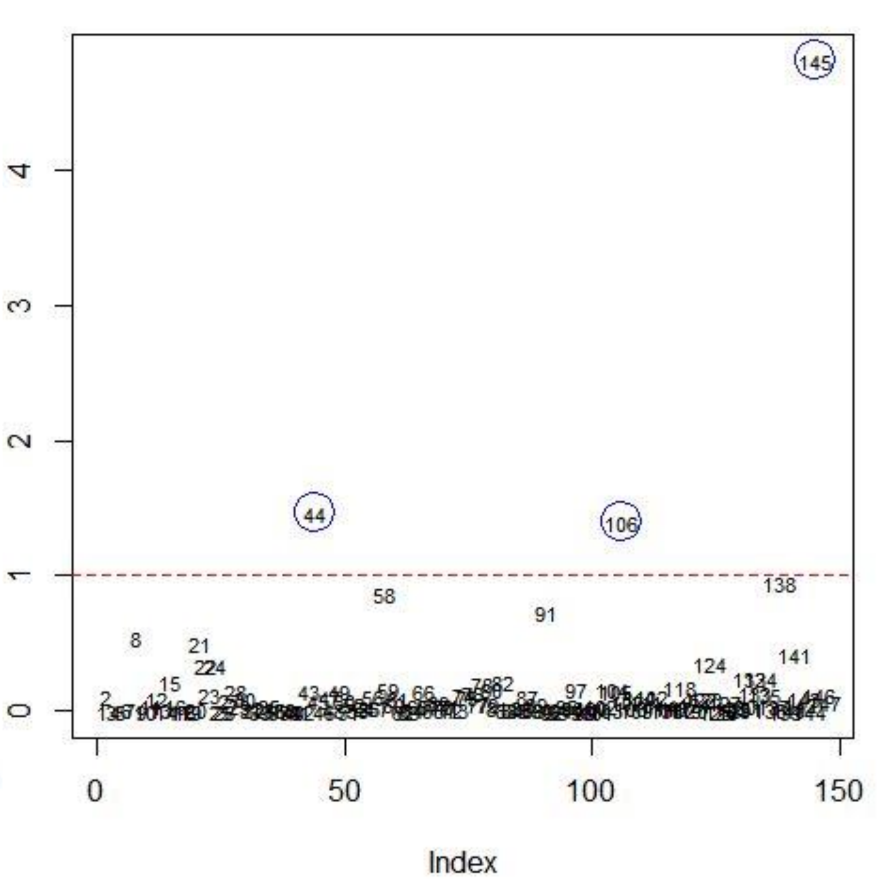
```
        Durbin-Watson test

data:  life ~ gdp + sani + pre + pri + sec + ter + smo + ob + al + co2 +
    hiv
DW = 2.2321, p-value = 0.9236
alternative hypothesis: true autocorrelation is greater than 0
```

02 데이터 분석_Outlier & Influential points

studentized residual

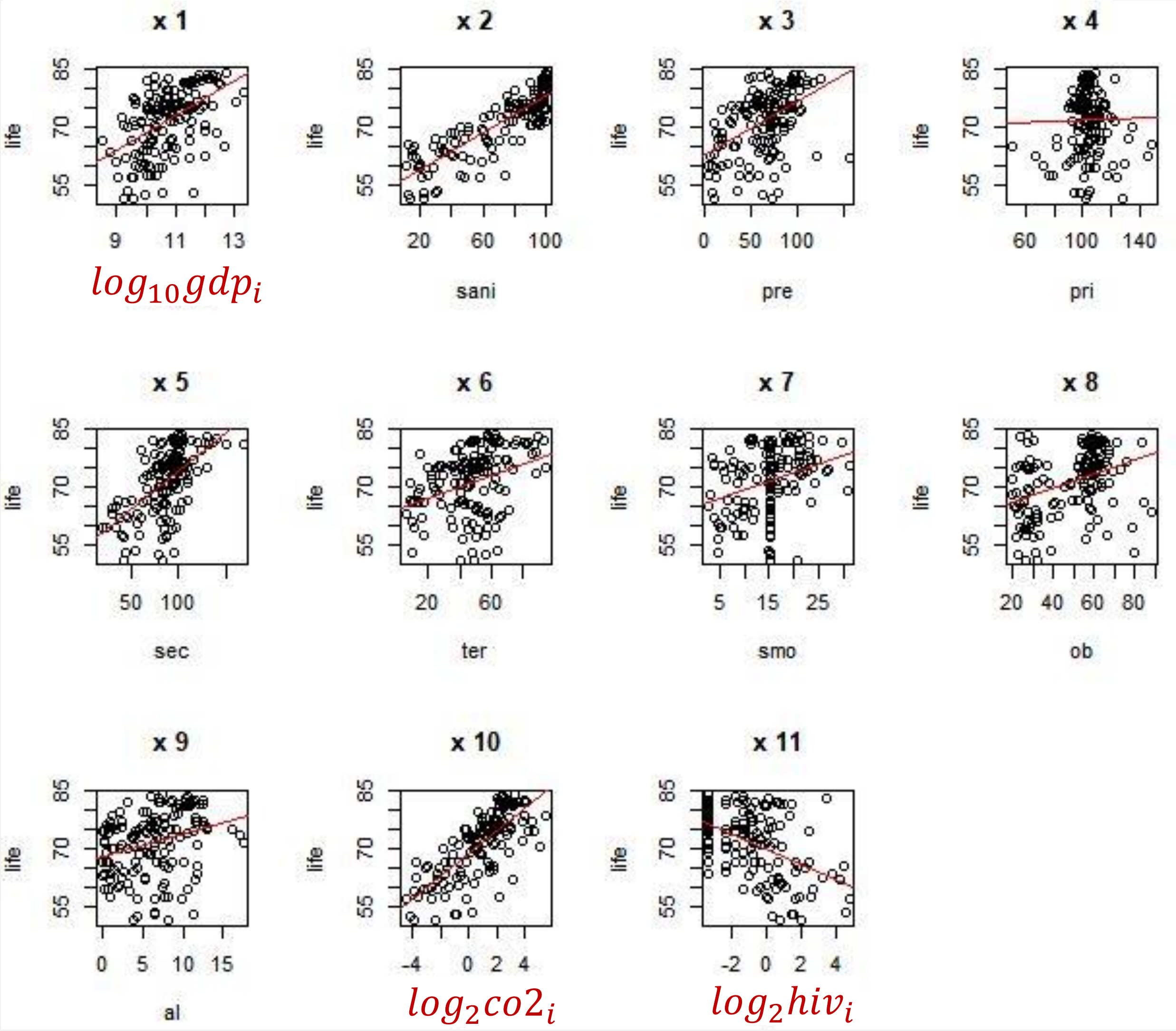


Cook's distance, influential point

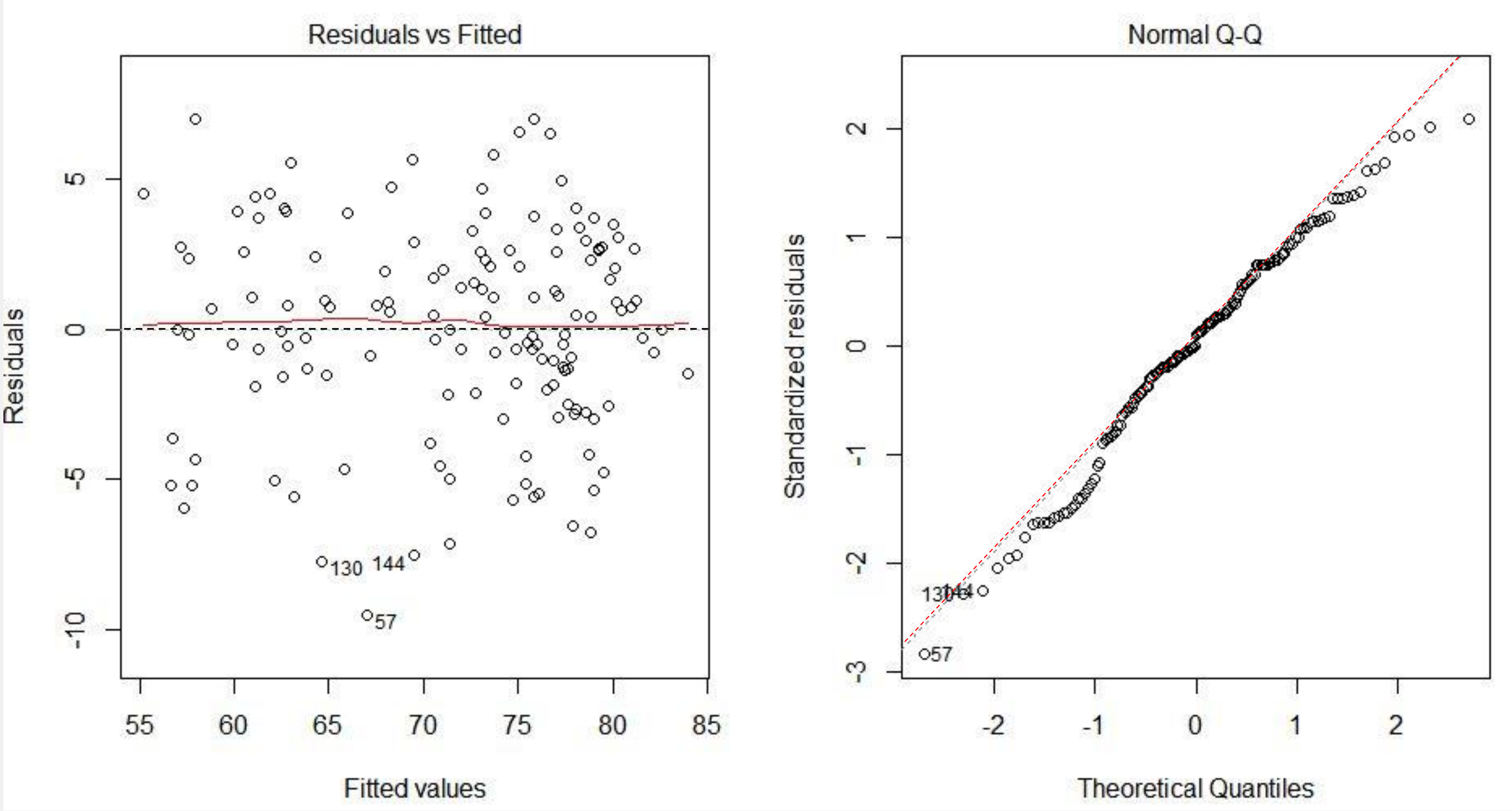


02 데이터 분석_Log transformation

Log transformation 후의 scatter plot

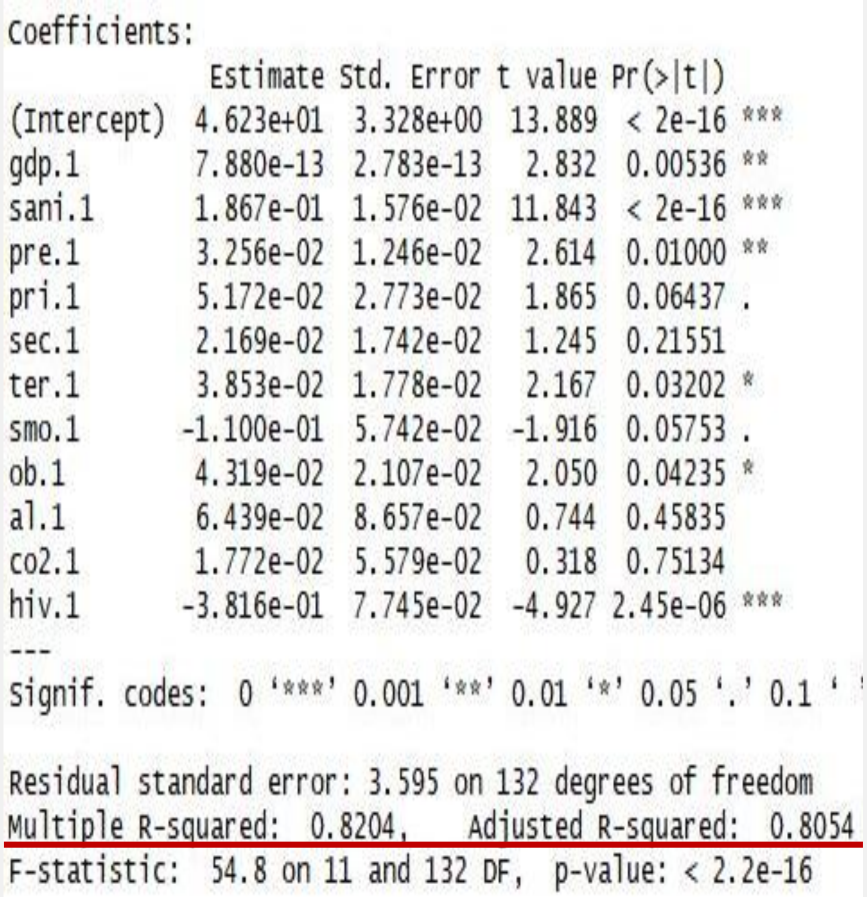


02 데이터 분석_Log transformation



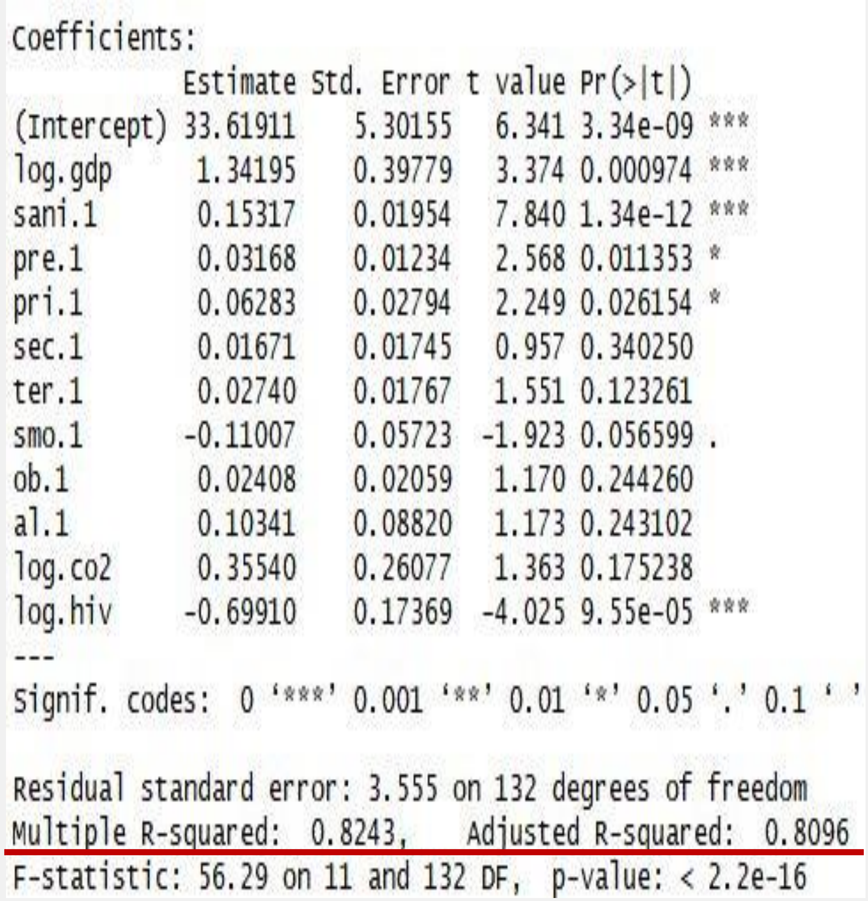
02 데이터 분석_Log transformation

기존 model



```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.623e+01  3.328e+00  13.889  < 2e-16 ***
gdp.1        7.880e-13  2.783e-13   2.832  0.00536 ** 
sani.1       1.867e-01  1.576e-02  11.843  < 2e-16 ***
pre.1        3.256e-02  1.246e-02   2.614  0.01000 ** 
pri.1        5.172e-02  2.773e-02   1.865  0.06437 .  
sec.1        2.169e-02  1.742e-02   1.245  0.21551    
ter.1        3.853e-02  1.778e-02   2.167  0.03202 *  
smo.1       -1.100e-01  5.742e-02  -1.916  0.05753 .  
ob.1         4.319e-02  2.107e-02   2.050  0.04235 *  
al.1         6.439e-02  8.657e-02   0.744  0.45835    
co2.1        1.772e-02  5.579e-02   0.318  0.75134    
hiv.1       -3.816e-01  7.745e-02  -4.927 2.45e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 3.595 on 132 degrees of freedom
Multiple R-squared:  0.8204,	Adjusted R-squared:  0.8054 
F-statistic:  54.8 on 11 and 132 DF,  p-value: < 2.2e-16
```

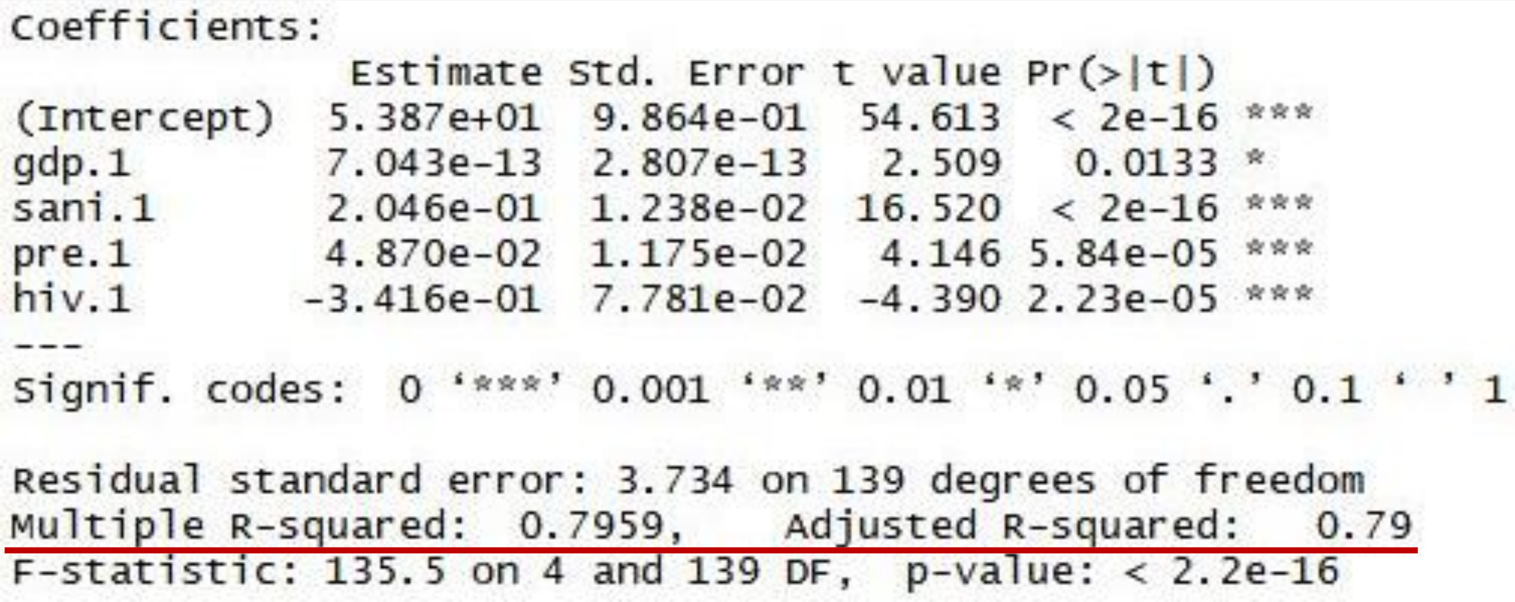
log transformation model



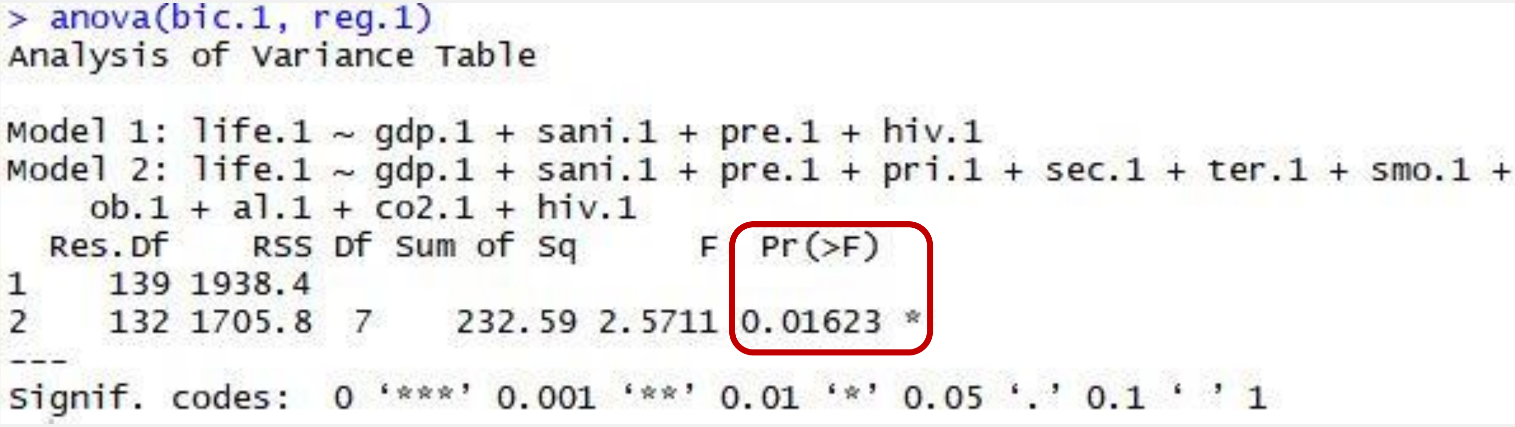
```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 33.61911    5.30155   6.341 3.34e-09 ***
log.gdp      1.34195    0.39779   3.374 0.000974 ***
sani.1       0.15317    0.01954   7.840 1.34e-12 ***
pre.1        0.03168    0.01234   2.568 0.011353 *  
pri.1        0.06283    0.02794   2.249 0.026154 *  
sec.1        0.01671    0.01745   0.957 0.340250    
ter.1        0.02740    0.01767   1.551 0.123261    
smo.1       -0.11007    0.05723  -1.923 0.056599 .  
ob.1         0.02408    0.02059   1.170 0.244260    
al.1         0.10341    0.08820   1.173 0.243102    
log.co2      0.35540    0.26077   1.363 0.175238    
log.hiv     -0.69910    0.17369  -4.025 9.55e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 3.555 on 132 degrees of freedom
Multiple R-squared:  0.8243,	Adjusted R-squared:  0.8096 
F-statistic: 56.29 on 11 and 132 DF,  p-value: < 2.2e-16
```

02 데이터 분석_Variable selection

life ~ gdp + sani + pre+ hiv



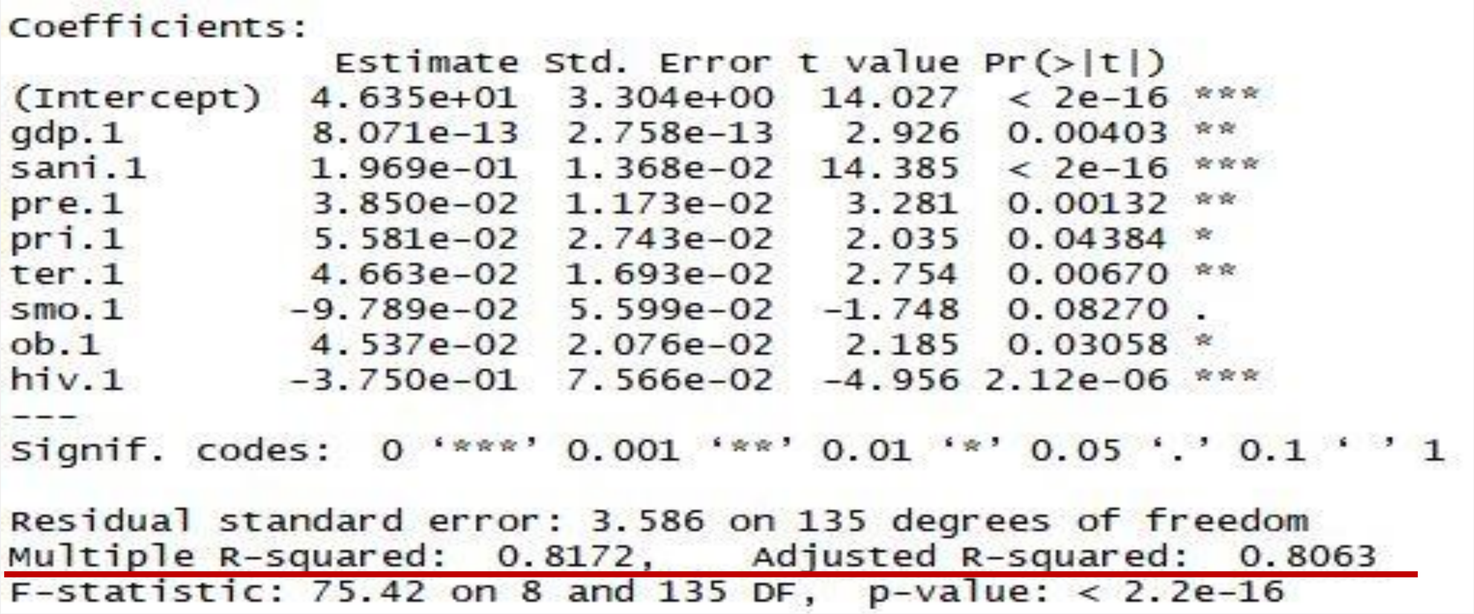
```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)   5.387e+01  9.864e-01  54.613  < 2e-16 ***
gdp.1         7.043e-13  2.807e-13   2.509   0.0133 *  
sani.1        2.046e-01  1.238e-02  16.520  < 2e-16 ***
pre.1         4.870e-02  1.175e-02   4.146 5.84e-05 ***
hiv.1        -3.416e-01  7.781e-02  -4.390 2.23e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.734 on 139 degrees of freedom
Multiple R-squared:  0.7959,	Adjusted R-squared:   0.79 
F-statistic: 135.5 on 4 and 139 DF,  p-value: < 2.2e-16
```



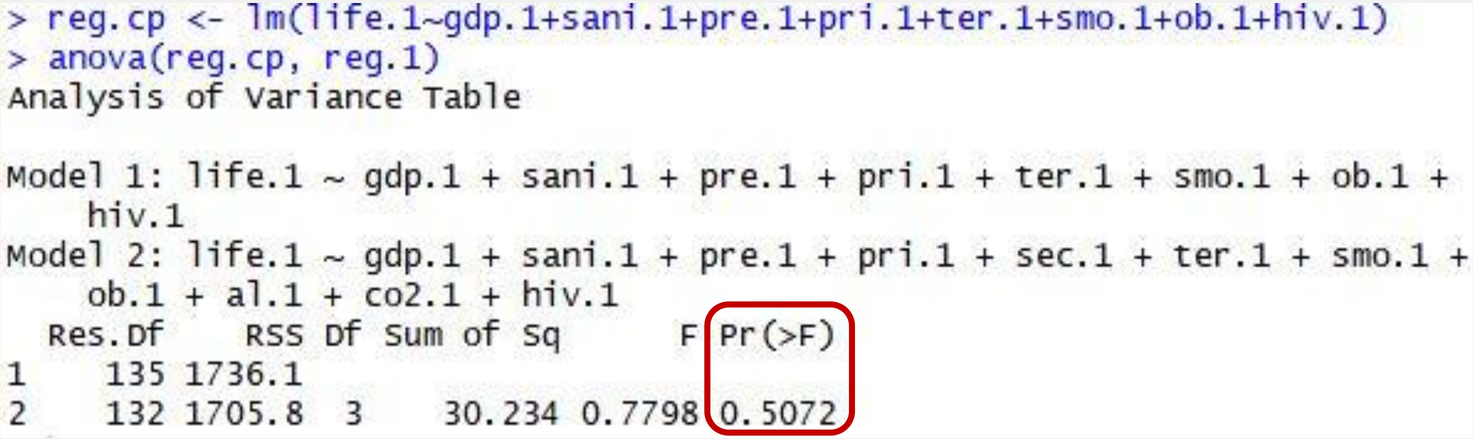
```
> anova(bic.1, reg.1)
Analysis of Variance Table

Model 1: life.1 ~ gdp.1 + sani.1 + pre.1 + hiv.1
Model 2: life.1 ~ gdp.1 + sani.1 + pre.1 + pri.1 + sec.1 + ter.1 + smo.1 + 
    ob.1 + al.1 + co2.1 + hiv.1
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)  
1    139 1938.4                              
2    132 1705.8  7    232.59 2.5711 0.01623 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

02 데이터 분석_Variable selection

life ~ gdp + sani + pre+ pri + ter + smo + ob + hiv



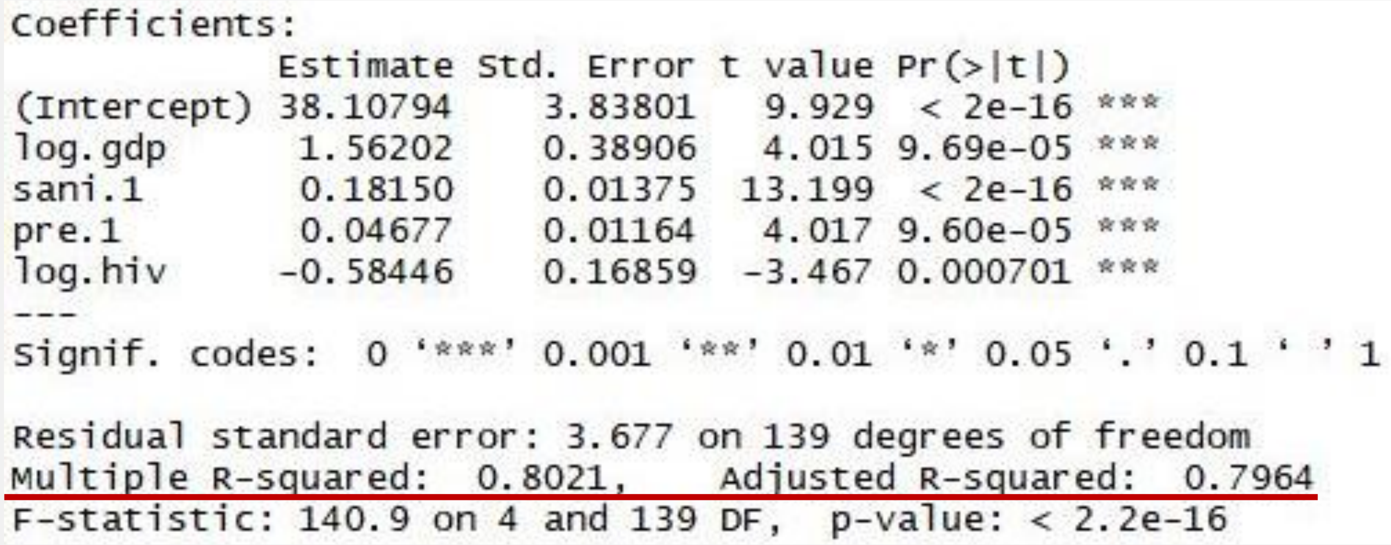
```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  4.635e+01  3.304e+00  14.027  < 2e-16 ***
gdp.1        8.071e-13  2.758e-13   2.926  0.00403 ** 
sani.1       1.969e-01  1.368e-02  14.385  < 2e-16 ***
pre.1        3.850e-02  1.173e-02   3.281  0.00132 ** 
pri.1        5.581e-02  2.743e-02   2.035  0.04384 *  
ter.1        4.663e-02  1.693e-02   2.754  0.00670 ** 
smo.1       -9.789e-02  5.599e-02  -1.748  0.08270 .  
ob.1         4.537e-02  2.076e-02   2.185  0.03058 *  
hiv.1       -3.750e-01  7.566e-02  -4.956 2.12e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.586 on 135 degrees of freedom
Multiple R-squared:  0.8172,	Adjusted R-squared:  0.8063 
F-statistic: 75.42 on 8 and 135 DF,  p-value: < 2.2e-16
```



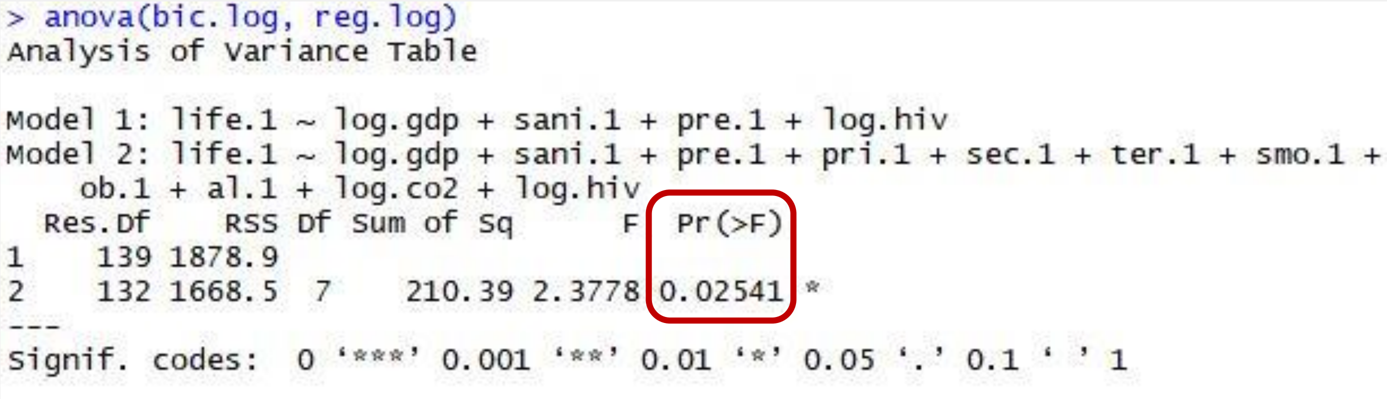
```
> reg.cp <- lm(life.1~gdp.1+sani.1+pre.1+pri.1+ter.1+smo.1+ob.1+hiv.1)
> anova(reg.cp, reg.1)
Analysis of Variance Table

Model 1: life.1 ~ gdp.1 + sani.1 + pre.1 + pri.1 + ter.1 + smo.1 + ob.1 + 
    hiv.1
Model 2: life.1 ~ gdp.1 + sani.1 + pre.1 + pri.1 + sec.1 + ter.1 + smo.1 + 
    ob.1 + al.1 + co2.1 + hiv.1
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    135 1736.1                           
2    132 1705.8  3    30.234 0.7798 0.5072
```

02 데이터 분석_Variable selection (log transformation)

life ~ $\log_{10}$gdp + sani + pre+ $\log_2$hiv



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 38.10794    3.83801   9.929  < 2e-16 ***
log.gdp      1.56202    0.38906   4.015 9.69e-05 ***
sani.1       0.18150    0.01375  13.199  < 2e-16 ***
pre.1        0.04677    0.01164   4.017 9.60e-05 ***
log.hiv     -0.58446    0.16859  -3.467 0.000701 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.677 on 139 degrees of freedom
Multiple R-squared:  0.8021,	Adjusted R-squared:  0.7964 
F-statistic: 140.9 on 4 and 139 DF,  p-value: < 2.2e-16
```



```
> anova(bic.log, reg.log)
Analysis of Variance Table

Model 1: life.1 ~ log.gdp + sani.1 + pre.1 + log.hiv
Model 2: life.1 ~ log.gdp + sani.1 + pre.1 + pri.1 + sec.1 + ter.1 + smo.1 + 
    ob.1 + al.1 + log.co2 + log.hiv
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)  
1    139 1878.9                              
2    132 1668.5  7    210.39 2.3778 0.02541 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

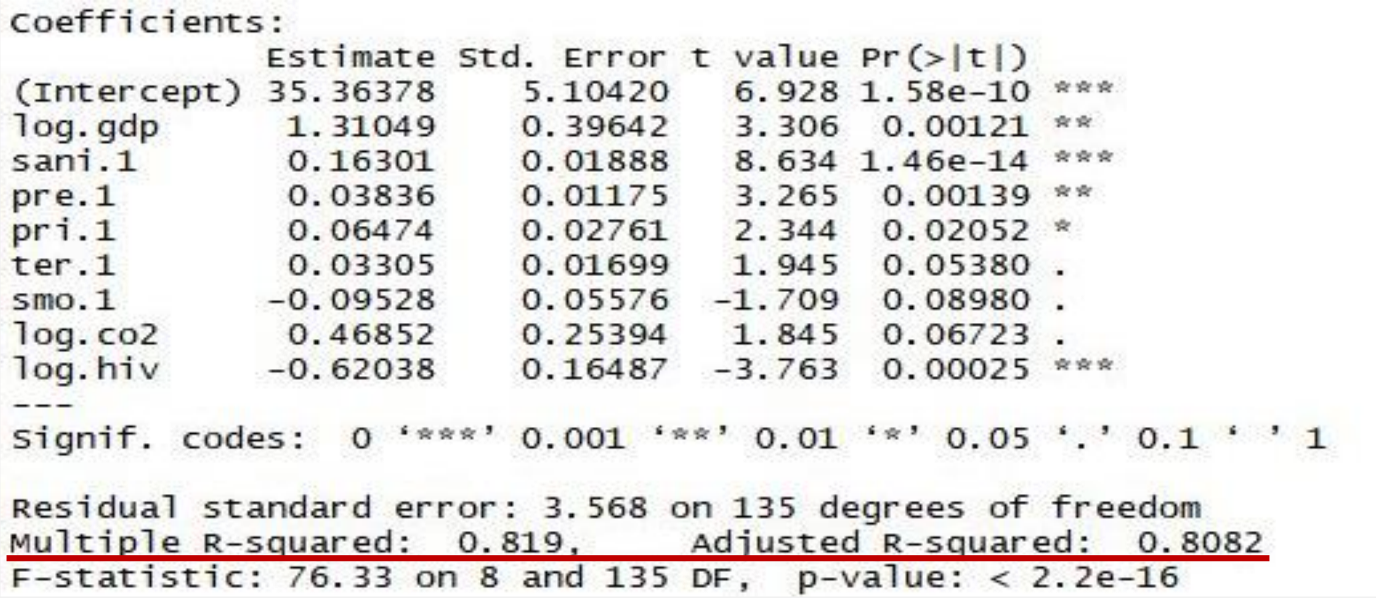
02 데이터 분석_Variable selection (log transformation)

01 전처리

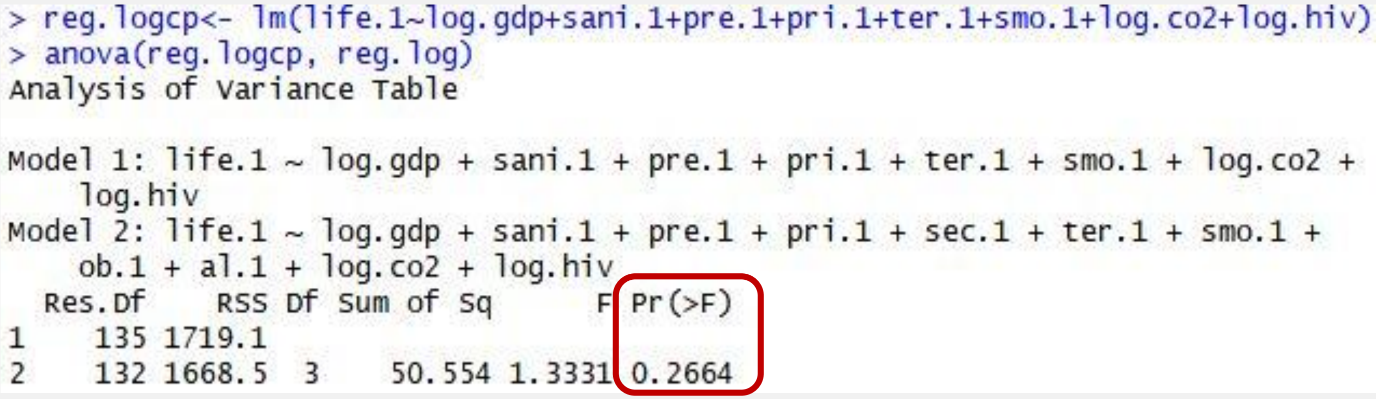
02 데이터 분석

03 한계점

life ~ $\log_{10}$gdp + sani + pre+ pri + ter + smo + $\log_2$co2 + $\log_2$hiv



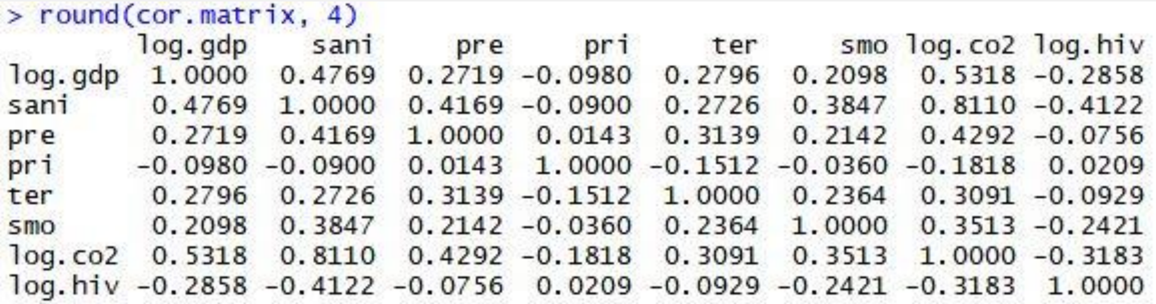
```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept) 35.36378    5.10420   6.928 1.58e-10 ***
log.gdp      1.31049    0.39642   3.306  0.00121 ** 
sani.1       0.16301    0.01888   8.634 1.46e-14 ***
pre.1        0.03836    0.01175   3.265  0.00139 ** 
pri.1        0.06474    0.02761   2.344  0.02052 *  
ter.1        0.03305    0.01699   1.945  0.05380 .  
smo.1       -0.09528    0.05576  -1.709  0.08980 .  
log.co2      0.46852    0.25394   1.845  0.06723 .  
log.hiv     -0.62038    0.16487  -3.763  0.00025 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.568 on 135 degrees of freedom
Multiple R-squared:  0.819,	Adjusted R-squared:  0.8082 
F-statistic: 76.33 on 8 and 135 DF,  p-value: < 2.2e-16
```



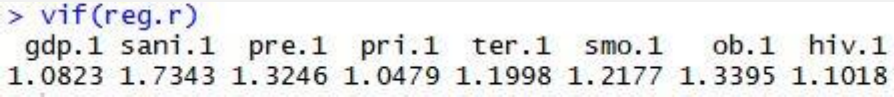
```
> reg.logcp<- lm(life.1~log.gdp+sani.1+pre.1+pri.1+ter.1+smo.1+log.co2+log.hiv)
> anova(reg.logcp, reg.log)
Analysis of Variance Table

Model 1: life.1 ~ log.gdp + sani.1 + pre.1 + pri.1 + ter.1 + smo.1 + log.co2 + 
    log.hiv
Model 2: life.1 ~ log.gdp + sani.1 + pre.1 + pri.1 + sec.1 + ter.1 + smo.1 + 
    ob.1 + al.1 + log.co2 + log.hiv
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1    135 1719.1                           
2    132 1668.5  3    50.554 1.3331 0.2664
```

02 데이터 분석_Multiple collinearity

<non log transformation reduced model>

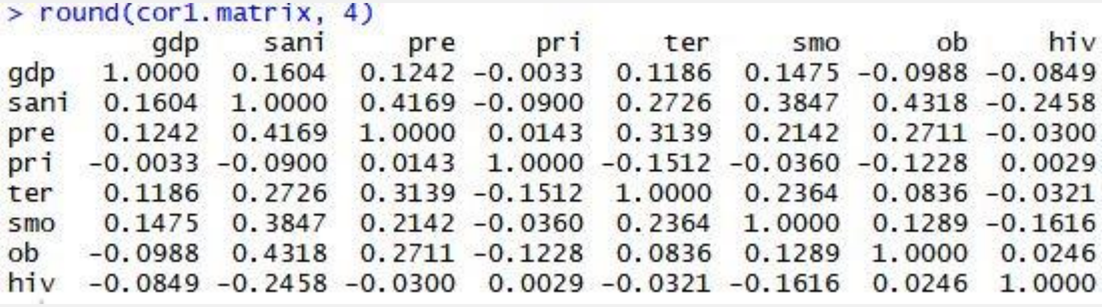


```
> round(cor.matrix, 4)
        log.gdp    sani     pre     pri     ter     smo log.co2 log.hiv
log.gdp  1.0000  0.4769  0.2719 -0.0980  0.2796  0.2098  0.5318 -0.2858
sani     0.4769  1.0000  0.4169 -0.0900  0.2726  0.3847  0.8110 -0.4122
pre      0.2719  0.4169  1.0000  0.0143  0.3139  0.2142  0.4292 -0.0756
pri     -0.0980 -0.0900  0.0143  1.0000 -0.1512 -0.0360 -0.1818  0.0209
ter      0.2796  0.2726  0.3139 -0.1512  1.0000  0.2364  0.3091 -0.0929
smo      0.2098  0.3847  0.2142 -0.0360  0.2364  1.0000  0.3513 -0.2421
log.co2  0.5318  0.8110  0.4292 -0.1818  0.3091  0.3513  1.0000 -0.3183
log.hiv -0.2858 -0.4122 -0.0756  0.0209 -0.0929 -0.2421 -0.3183  1.0000
```

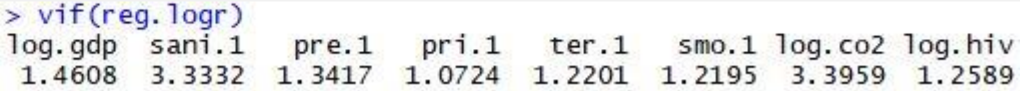


```
> vif(reg.r)
  gdp.1  sani.1   pre.1   pri.1   ter.1   smo.1    ob.1   hiv.1
1.0823 1.7343 1.3246 1.0479 1.1998 1.2177 1.3395 1.1018
```

<log transformation reduced model>

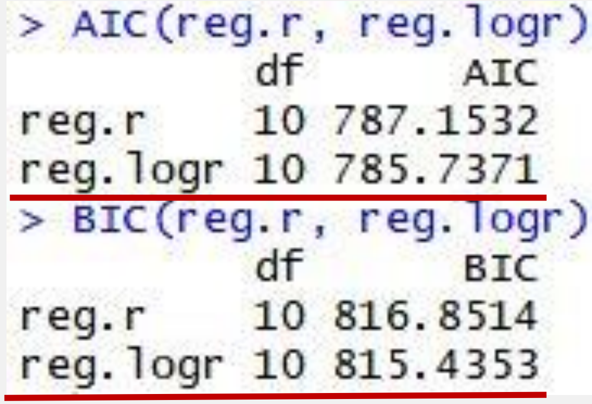


```
> round(cor1.matrix, 4)
         gdp    sani     pre     pri     ter     smo      ob     hiv
gdp   1.0000  0.1604  0.1242 -0.0033  0.1186  0.1475 -0.0988 -0.0849
sani  0.1604  1.0000  0.4169 -0.0900  0.2726  0.3847  0.4318 -0.2458
pre   0.1242  0.4169  1.0000  0.0143  0.3139  0.2142  0.2711 -0.0300
pri  -0.0033 -0.0900  0.0143  1.0000 -0.1512 -0.0360 -0.1228  0.0029
ter   0.1186  0.2726  0.3139 -0.1512  1.0000  0.2364  0.0836 -0.0321
smo   0.1475  0.3847  0.2142 -0.0360  0.2364  1.0000  0.1289 -0.1616
ob   -0.0988  0.4318  0.2711 -0.1228  0.0836  0.1289  1.0000  0.0246
hiv  -0.0849 -0.2458 -0.0300  0.0029 -0.0321 -0.1616  0.0246  1.0000
```

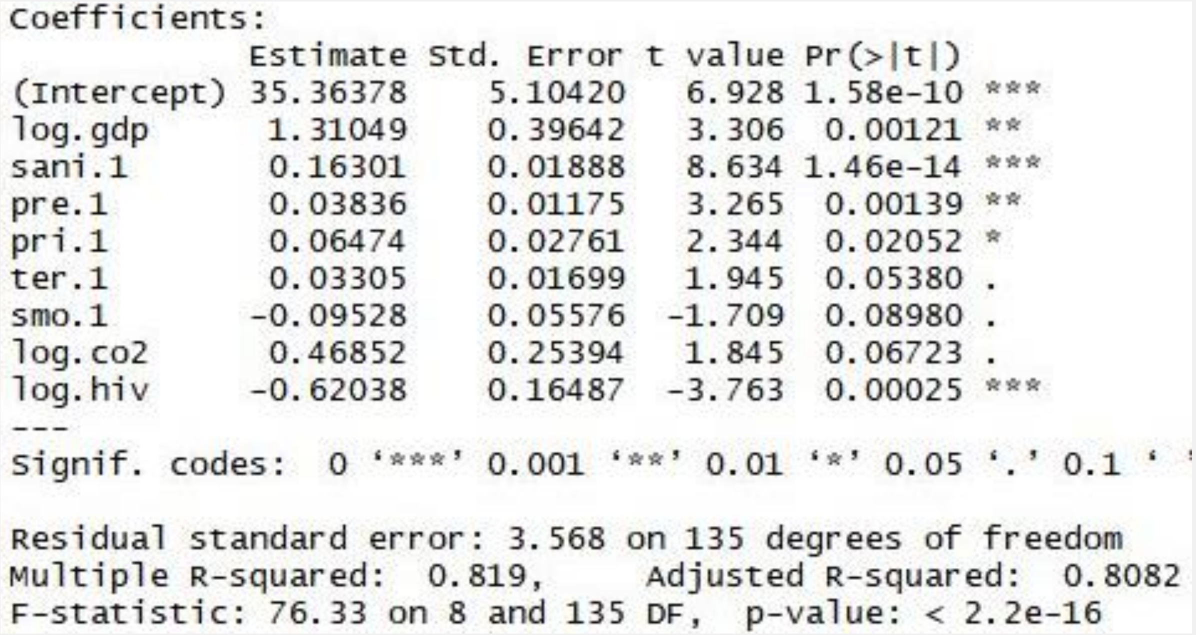


```
> vif(reg.logr)
log.gdp  sani.1   pre.1   pri.1   ter.1   smo.1 log.co2 log.hiv
 1.4608  3.3332  1.3417  1.0724  1.2201  1.2195  3.3959  1.2589
```

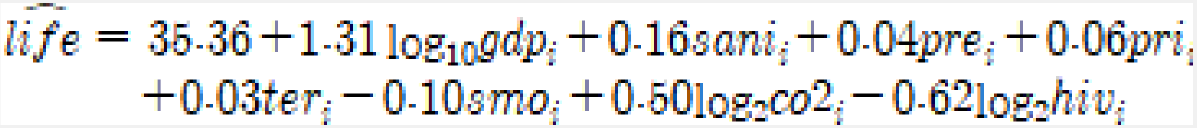
02 데이터 분석_Model selection



```
> AIC(reg.r, reg.logr)
          df      AIC
reg.r     10 787.1532
reg.logr  10 785.7371
> BIC(reg.r, reg.logr)
          df      BIC
reg.r     10 816.8514
reg.logr  10 815.4353
```



```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 35.36378    5.10420   6.928 1.58e-10 ***
log.gdp      1.31049    0.39642   3.306  0.00121 ** 
sani.1       0.16301    0.01888   8.634 1.46e-14 ***
pre.1        0.03836    0.01175   3.265  0.00139 ** 
pri.1        0.06474    0.02761   2.344  0.02052 *  
ter.1        0.03305    0.01699   1.945  0.05380 .  
smo.1       -0.09528    0.05576  -1.709  0.08980 .  
log.co2      0.46852    0.25394   1.845  0.06723 .  
log.hiv     -0.62038    0.16487  -3.763  0.00025 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’

Residual standard error: 3.568 on 135 degrees of freedom
Multiple R-squared:  0.819,	Adjusted R-squared:  0.8082 
F-statistic: 76.33 on 8 and 135 DF,  p-value: < 2.2e-16
```



$$\widehat{life} = 35.36 + 1.31\log_{10}gdp_i + 0.16sani_i + 0.04pre_i + 0.06pri_i \\ + 0.03ter_i - 0.10smo_i + 0.50\log_2 co2_i - 0.62\log_2 hiv_i$$

01 전처리
02 데이터 분석
03 한계점

03 한계점_

시점 통일 불가능.

Log transformation 순서

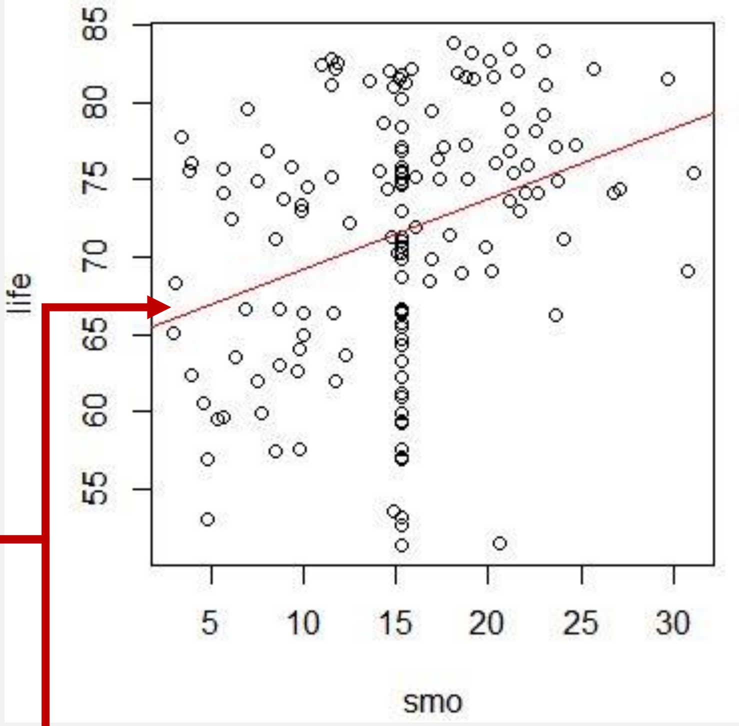
Heteroscedasticity 처리 – WLS

BIC, Mallow Cp 차이

Obesity cluster

Smoking의 beta coefficient 차이

기대 수명과 사회적 요건 간의 관계 해석



$$\widehat{life} = 35.36 + 1.31\log_{10}gdp_i + 0.16sani_i + 0.04pre_i + 0.06pri_i + 0.03ter_i \boxed{-0.10smo_i} + 0.50\log_2 co2_i - 0.62\log_2 hiv_i$$

Thank You

1. 김명중, 박범조. R을 이용한 분위수회귀 분석 : 경제외적 요인이 기대수명에 미치는 영향,
DKU 미래산업연구소_ 단국대학교 산업연구 37권 2호, 2013, p33-68.
2. 최용옥, 급속한 기대수명 증가의 함의(Longevity Risk in Korea), KDI FOCUS, NO.69, (Korea Development Institute), 2016, p3
3. https://data.worldbank.org
4. https://data.humdata.org/dataset/prevalence_of_hiv_total_of_population_aged_15-49/resource/c5f56338-471b-4aaf-b5b9-b1f7db160bc1
5. http://www.datamarket.kr/xe/board_BoGi29/9880
6. http://www.saedsayad.com/k_nearest_neighbors_reg.htm