Colony.R

KIM

2019-06-21

library(readxl); library(car); library(psych); library(DMwR)

## Loading required package: carData

##   
## Attaching package: 'psych'

## The following object is masked from 'package:car':  
##   
## logit

## Loading required package: lattice

## Loading required package: grid

## Registered S3 method overwritten by 'xts':  
## method from  
## as.zoo.xts zoo

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

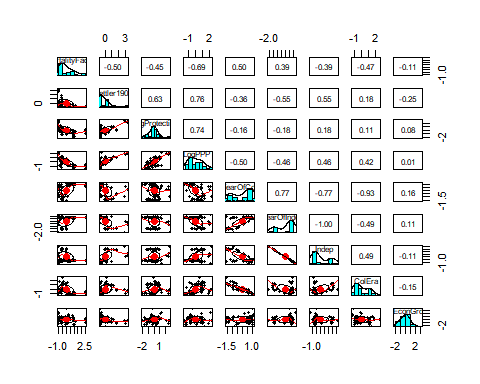
Raw0 <- read.csv("Rawdata\_2019.csv")  
str(Raw0)

## 'data.frame': 63 obs. of 10 variables:  
## $ Country : Factor w/ 63 levels "ARG","AUS","BEN",..: 1 2 3 4 5 6 7 8 9 10 ...  
## $ MortalityFactor: num 68.9 8.55 NA 280 71.41 ...  
## $ Settler1900 : num 0.6 0.98 NA NA 0 NA 0.4 0.2 0.99 0.5 ...  
## $ LogProtection : num 6.39 9.32 NA 4.45 5.14 NA 7.91 NA 9.73 7.82 ...  
## $ LogPPP : num 9.13 9.9 NA 6.85 6.88 NA 8.73 9.27 9.99 9.34 ...  
## $ YearOfCol : int 1650 1770 1894 1896 1757 1535 1531 1627 1763 1557 ...  
## $ YearOfIndep : int 1816 1901 1960 1960 1971 1825 1822 1966 1867 1818 ...  
## $ Indep : int 184 99 40 40 29 175 178 34 133 182 ...  
## $ ColEra : int 166 131 66 64 214 290 291 339 104 261 ...  
## $ AvgEconGrowth : num 2.59 3.71 3.37 3.81 3.45 ...

KNN0 <- knnImputation(Raw0[, !names(Raw0) %in% "medv"])  
anyNA(KNN0);

## [1] FALSE

KNN0.1 <- as.data.frame(scale(KNN0[2:10]))   
  
Sca\_temp <- as.data.frame(scale(KNN0[2:10]))  
Sca0 <- cbind(Raw0$Country, Sca\_temp) # 정규화  
  
pairs.panels(Sca\_temp)



SM1 <- lm(Settler1900 ~ MortalityFactor, data = Sca0)  
summary(SM1)

##   
## Call:  
## lm(formula = Settler1900 ~ MortalityFactor, data = Sca0)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.2106 -0.4814 -0.1883 0.1818 2.7138   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2.168e-16 1.101e-01 0.000 1   
## MortalityFactor -4.989e-01 1.110e-01 -4.496 3.15e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.8737 on 61 degrees of freedom  
## Multiple R-squared: 0.2489, Adjusted R-squared: 0.2366   
## F-statistic: 20.22 on 1 and 61 DF, p-value: 3.155e-05

SM2 <- lm(LogProtection ~ Settler1900, data = Sca0)  
summary(SM2)

##   
## Call:  
## lm(formula = LogProtection ~ Settler1900, data = Sca0)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.6312 -0.5306 0.1002 0.4772 1.7849   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.697e-16 9.829e-02 0.000 1   
## Settler1900 6.334e-01 9.908e-02 6.393 2.53e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7801 on 61 degrees of freedom  
## Multiple R-squared: 0.4012, Adjusted R-squared: 0.3914   
## F-statistic: 40.88 on 1 and 61 DF, p-value: 2.527e-08

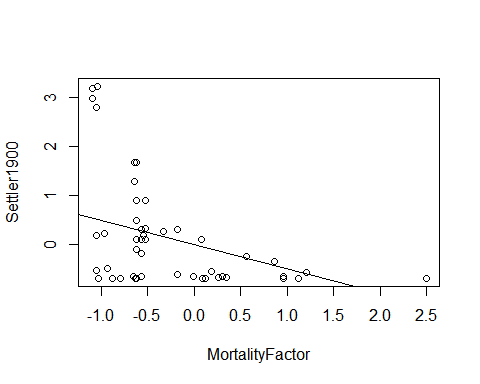
SM3 <- lm(LogPPP ~ LogProtection, data = Sca0)  
summary(SM3)

##   
## Call:  
## lm(formula = LogPPP ~ LogProtection, data = Sca0)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.82489 -0.43152 -0.06341 0.48501 1.22522   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.270e-16 8.590e-02 0.000 1   
## LogProtection 7.366e-01 8.659e-02 8.507 5.93e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.6818 on 61 degrees of freedom  
## Multiple R-squared: 0.5426, Adjusted R-squared: 0.5351   
## F-statistic: 72.37 on 1 and 61 DF, p-value: 5.925e-12

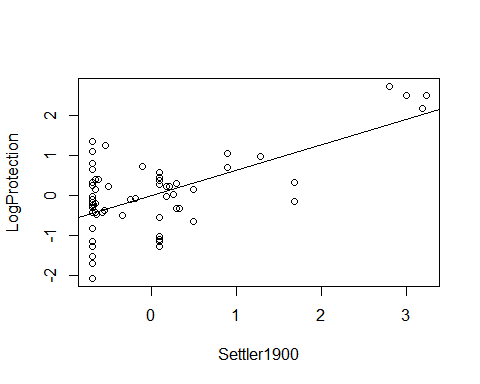
SM4 <- lm(LogPPP ~ MortalityFactor, data = Sca0)  
summary(SM4)

##   
## Call:  
## lm(formula = LogPPP ~ MortalityFactor, data = Sca0)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.75243 -0.55852 -0.03117 0.42106 1.55499   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.573e-16 9.149e-02 0.000 1   
## MortalityFactor -6.937e-01 9.222e-02 -7.522 2.93e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7262 on 61 degrees of freedom  
## Multiple R-squared: 0.4812, Adjusted R-squared: 0.4727   
## F-statistic: 56.58 on 1 and 61 DF, p-value: 2.932e-10

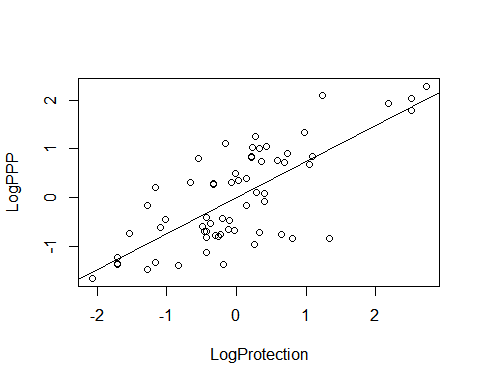
with(Sca0, plot(MortalityFactor, Settler1900))  
abline(SM1)



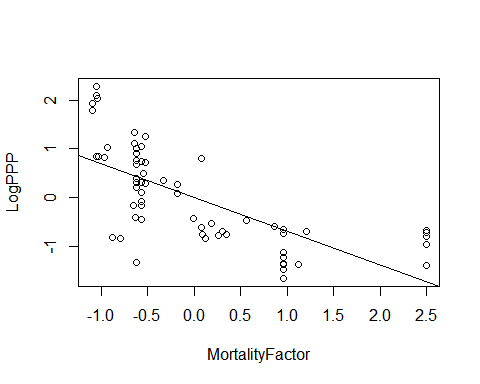
with(Sca0, plot(Settler1900, LogProtection))  
abline(SM2)



with(Sca0, plot(LogProtection, LogPPP))  
abline(SM3)



with(Sca0, plot(MortalityFactor, LogPPP))  
abline(SM4)



Model <- lm(LogPPP ~ . -`Raw0$Country`, data = Sca0)  
Model.step <- step(Model, direction = "both")

## Start: AIC=-93.31  
## LogPPP ~ (`Raw0$Country` + MortalityFactor + Settler1900 + LogProtection +   
## YearOfCol + YearOfIndep + Indep + ColEra + AvgEconGrowth) -   
## `Raw0$Country`  
##   
##   
## Step: AIC=-93.31  
## LogPPP ~ MortalityFactor + Settler1900 + LogProtection + YearOfCol +   
## YearOfIndep + Indep + AvgEconGrowth  
##   
##   
## Step: AIC=-93.31  
## LogPPP ~ MortalityFactor + Settler1900 + LogProtection + YearOfCol +   
## YearOfIndep + AvgEconGrowth  
##   
## Df Sum of Sq RSS AIC  
## <none> 11.470 -93.314  
## - YearOfIndep 1 0.3939 11.864 -93.187  
## - AvgEconGrowth 1 0.4085 11.878 -93.110  
## - MortalityFactor 1 1.5375 13.008 -87.389  
## - YearOfCol 1 1.8063 13.276 -86.101  
## - Settler1900 1 3.3108 14.781 -79.338  
## - LogProtection 1 3.3210 14.791 -79.294

confint(Model.step)

## 2.5 % 97.5 %  
## (Intercept) -0.11422195 0.1142219  
## MortalityFactor -0.37999372 -0.0590139  
## Settler1900 0.21670399 0.6471036  
## LogProtection 0.16901397 0.5036617  
## YearOfCol -0.52659172 -0.1023365  
## YearOfIndep -0.06868760 0.3776644  
## AvgEconGrowth -0.04019291 0.2322734

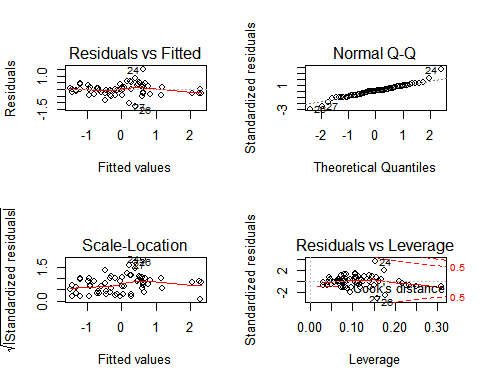
summary(Model.step)

##   
## Call:  
## lm(formula = LogPPP ~ MortalityFactor + Settler1900 + LogProtection +   
## YearOfCol + YearOfIndep + AvgEconGrowth, data = Sca0)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.24049 -0.26542 0.00245 0.21187 1.48139   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.245e-16 5.702e-02 0.000 1.000000   
## MortalityFactor -2.195e-01 8.012e-02 -2.740 0.008231 \*\*   
## Settler1900 4.319e-01 1.074e-01 4.020 0.000175 \*\*\*  
## LogProtection 3.363e-01 8.353e-02 4.027 0.000172 \*\*\*  
## YearOfCol -3.145e-01 1.059e-01 -2.970 0.004383 \*\*   
## YearOfIndep 1.545e-01 1.114e-01 1.387 0.171030   
## AvgEconGrowth 9.604e-02 6.801e-02 1.412 0.163417   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.4526 on 56 degrees of freedom  
## Multiple R-squared: 0.815, Adjusted R-squared: 0.7952   
## F-statistic: 41.12 on 6 and 56 DF, p-value: < 2.2e-16

summary.aov(Model.step)

## Df Sum Sq Mean Sq F value Pr(>F)   
## MortalityFactor 1 29.833 29.833 145.656 < 2e-16 \*\*\*  
## Settler1900 1 14.165 14.165 69.156 2.34e-11 \*\*\*  
## LogProtection 1 4.311 4.311 21.046 2.57e-05 \*\*\*  
## YearOfCol 1 1.596 1.596 7.791 0.00717 \*\*   
## YearOfIndep 1 0.217 0.217 1.061 0.30743   
## AvgEconGrowth 1 0.408 0.408 1.994 0.16342   
## Residuals 56 11.470 0.205   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

par(mfrow = c(2,2))  
plot(Model.step)



vif(Model.step)

## MortalityFactor Settler1900 LogProtection YearOfCol   
## 1.942887 3.493299 2.111874 3.394270   
## YearOfIndep AvgEconGrowth   
## 3.757051 1.399966