HW_MLR

경제학과 2020110210 공소연

2022-10-11

Q1: Check the first 3 rows of this dataset.

```
state.df <- data.frame(state.x77)

#Q1
head(state.df, 3)
```

```
##         Population Income Illiteracy Life.Exp Murder HS.Grad Frost
Area
## Alabama       3615   3624        2.1    69.05   15.1    41.3    20
50708
## Alaska         365   6315        1.5    69.31   11.3    66.7   152
566432
## Arizona       2212   4530        1.8    70.55    7.8    58.1    15
113417
```

Q2: Create a new data frame using variable “Murder”, “Population”, “Illiteracy”, “Income” and “Frost” and check it.

```
selected.var <- c(5,1,3,2,7)
stat.df <- state.df[selected.var]
str(stat.df)
```

```
## 'data.frame':    50 obs. of  5 variables:
##  $ Murder    : num  15.1 11.3 7.8 10.1 10.3 6.8 3.1 6.2 10.7
13.9 ...
##  $ Population: num  3615 365 2212 2110 21198 ...
##  $ Illiteracy: num  2.1 1.5 1.8 1.9 1.1 0.7 1.1 0.9 1.3 2 ...
##  $ Income    : num  3624 6315 4530 3378 5114 ...
##  $ Frost     : num  20 152 15 65 20 166 139 103 11 60 ...
```

Q3: Conduct a multiple regression to predict the dependent variable “Murder” using all other variables as the independent variables and show the results.

```
# partition data
set.seed(1)
train.index <- sample(c(1:50), 30)
train.df <- stat.df[train.index,]
valid.df <- stat.df[-train.index,]

stat.lm <- lm(Murder~., data = train.df)
options(scipen =999)
summary(stat.lm)
```

```
##
## Call:
## lm(formula = Murder ~ ., data = train.df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.828 -1.333 -0.321  1.038  3.708
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.4403595  5.6851019   2.188  0.03821 *
## Population   0.0005081  0.0001803   2.817  0.00932 **
## Illiteracy   2.3582868  1.0603049   2.224  0.03540 *
## Income      -0.0019100  0.0010688  -1.787  0.08607 .
## Frost       -0.0172656  0.0118441  -1.458  0.15736
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.197 on 25 degrees of freedom
## Multiple R-squared:  0.7523, Adjusted R-squared:  0.7127
## F-statistic: 18.98 on 4 and 25 DF,  p-value: 0.000000276
```

Q4: Use stepwise method to select independent variables and conduct multiple regression with the selected independent variables.

```
stat.lm.step <- step(stat.lm, direction = "both")
```

```
## Start:  AIC=51.75
## Murder ~ Population + Illiteracy + Income + Frost
##
##              Df Sum of Sq    RSS    AIC
## <none>                     120.63 51.746
## - Frost       1    10.254 130.89 52.194
## - Income      1    15.409 136.04 53.353
## - Illiteracy  1    23.870 144.50 55.163
## - Population  1    38.301 158.93 58.018
```

```
summary(stat.lm.step)
```

```
##
## Call:
## lm(formula = Murder ~ Population + Illiteracy + Income + Frost,
##     data = train.df)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.828 -1.333 -0.321  1.038  3.708
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.4403595  5.6851019   2.188  0.03821 *
## Population   0.0005081  0.0001803   2.817  0.00932 **
```

```
## Illiteracy   2.3582868  1.0603049   2.224  0.03540 *
## Income      -0.0019100  0.0010688  -1.787  0.08607 .
## Frost       -0.0172656  0.0118441  -1.458  0.15736
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.197 on 25 degrees of freedom
## Multiple R-squared:  0.7523, Adjusted R-squared:  0.7127
## F-statistic: 18.98 on 4 and 25 DF,  p-value: 0.000000276
```

Q5: Compare the results in Q3 and Q4.

```
coef(stat.lm)
```

```
##   (Intercept)    Population    Illiteracy        Income
Frost
## 12.4403595341  0.0005080746  2.3582867626 -0.0019100069 -
0.0172655704
```

```
coef(stat.lm.step)
```

```
##   (Intercept)    Population    Illiteracy        Income
Frost
## 12.4403595341  0.0005080746  2.3582867626 -0.0019100069 -
0.0172655704
```

```
stat.lm.pred <- predict(stat.lm, valid.df)
stat.lm.step.pred <- predict(stat.lm.step, valid.df)

# 모델의 평가측도 계산
library(forecast)
```

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

```
accuracy(stat.lm.pred, valid.df$Murder)
```

```
##                 ME     RMSE      MAE      MPE    MAPE
## Test set 0.7422927 3.990812 2.815258 2.608088 35.7761
```

```
accuracy(stat.lm.step.pred, valid.df$Murder)
```

```
##                 ME     RMSE      MAE      MPE    MAPE
## Test set 0.7422927 3.990812 2.815258 2.608088 35.7761
```

```
data.frame("Predicted" = stat.lm.pred,
           "Actual" = valid.df$Murder,
           "Residual" = valid.df$Murder - stat.lm.pred)
```

```
##               Predicted Actual     Residual
## Alaska        1.4771766   11.3  9.822823427
## California   15.6915528   10.3 -5.391552836
## Delaware      3.8934158    6.2  2.306584177
```

```
## Hawaii         7.8827488    6.2 -1.682748818
## Illinois       8.3045958   10.3  1.995404208
## Kansas         4.1276443    4.5  0.372355680
## Kentucky       9.2042921   10.6  1.395707927
## Maine          4.7933808    2.7 -2.093380808
## Mississippi   12.5091704   12.5 -0.009170398
## Nevada         0.8387141   11.5 10.661285891
## New Jersey     6.7719390    5.2 -1.571938977
## New York      14.1453762   10.9 -3.245376173
## Ohio           8.9286972    7.4 -1.528697151
## Oregon         5.3554566    4.2 -1.155456606
## Pennsylvania 10.1513280    6.1 -4.051328020
## Tennessee     10.0629159   11.0  0.937084138
## Utah           4.4191144    4.5  0.080885633
## Vermont        3.7321300    5.5  1.767870026
## Virginia       7.8261644    9.5  1.673835554
## Wyoming        2.3383324    6.9  4.561667589
```

```
data.frame("Predicted" = stat.lm.step.pred,
           "Actual" = valid.df$Murder,
           "Residual" = valid.df$Murder - stat.lm.step.pred)
```

```
##               Predicted Actual     Residual
## Alaska         1.4771766   11.3  9.822823427
## California    15.6915528   10.3 -5.391552836
## Delaware       3.8934158    6.2  2.306584177
## Hawaii         7.8827488    6.2 -1.682748818
## Illinois       8.3045958   10.3  1.995404208
## Kansas         4.1276443    4.5  0.372355680
## Kentucky       9.2042921   10.6  1.395707927
## Maine          4.7933808    2.7 -2.093380808
## Mississippi   12.5091704   12.5 -0.009170398
## Nevada         0.8387141   11.5 10.661285891
## New Jersey     6.7719390    5.2 -1.571938977
## New York      14.1453762   10.9 -3.245376173
## Ohio           8.9286972    7.4 -1.528697151
## Oregon         5.3554566    4.2 -1.155456606
## Pennsylvania 10.1513280    6.1 -4.051328020
## Tennessee     10.0629159   11.0  0.937084138
## Utah           4.4191144    4.5  0.080885633
## Vermont        3.7321300    5.5  1.767870026
## Virginia       7.8261644    9.5  1.673835554
## Wyoming        2.3383324    6.9  4.561667589
```