# IIMT2641 Assignment 3

Sibo Ding

Spring 2023

https://rpubs.com/aashish/ClimateChangeDataAnalysis

## Load the Data

```
climate_change <- read.csv("ClimateChange.csv")
head(climate_change) # First 6 rows
```

```
##   Year Month   MEI    CO2     CH4     N2O  CFC.11  CFC.12      TSI
Aerosols
## 1 1983     5 2.556 345.96 1638.59 303.677 191.324 350.113 1366.102
0.0863
## 2 1983     6 2.167 345.52 1633.71 303.746 192.057 351.848 1366.121
0.0794
## 3 1983     7 1.741 344.15 1633.22 303.795 192.818 353.725 1366.285
0.0731
## 4 1983     8 1.130 342.25 1631.35 303.839 193.602 355.633 1366.420
0.0673
## 5 1983     9 0.428 340.17 1648.40 303.901 194.392 357.465 1366.234
0.0619
## 6 1983    10 0.002 340.30 1663.79 303.970 195.171 359.174 1366.059
0.0569
##    Temp
## 1 0.109
## 2 0.118
## 3 0.137
## 4 0.176
## 5 0.149
## 6 0.093
```

```
dim(climate_change) # Number of observations and variables
```

```
## [1] 308  11
```

```
names(climate_change) # Names of variables
```

```
##  [1] "Year"     "Month"    "MEI"      "CO2"      "CH4"      "N2O"
##  [7] "CFC.11"   "CFC.12"   "TSI"      "Aerosols" "Temp"
```

## Train-test Split

```
climate_train <- climate_change |> subset(Year <= 2006)
climate_test <- climate_change |> subset(Year > 2006)
```

## Build Linear Regression Model

```
climate_reg1 <- lm(Temp ~ . - Year - Month, data = climate_train)
summary(climate_reg1)
```

```
##
## Call:
## lm(formula = Temp ~ . - Year - Month, data = climate_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.25888 -0.05913 -0.00082  0.05649  0.32433
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.246e+02  1.989e+01  -6.265 1.43e-09 ***
## MEI          6.421e-02  6.470e-03   9.923  < 2e-16 ***
## CO2          6.457e-03  2.285e-03   2.826  0.00505 **
## CH4          1.240e-04  5.158e-04   0.240  0.81015
## N2O         -1.653e-02  8.565e-03  -1.930  0.05467 .
## CFC.11      -6.631e-03  1.626e-03  -4.078 5.96e-05 ***
## CFC.12       3.808e-03  1.014e-03   3.757  0.00021 ***
## TSI          9.314e-02  1.475e-02   6.313 1.10e-09 ***
## Aerosols    -1.538e+00  2.133e-01  -7.210 5.41e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09171 on 275 degrees of freedom
## Multiple R-squared:  0.7509, Adjusted R-squared:  0.7436
## F-statistic: 103.6 on 8 and 275 DF,  p-value: < 2.2e-16
```

```
# R^2
summary(climate_reg1)$r.squared
```

```
## [1] 0.7508933
```

```
# Significant independent variables (with p < 0.05)
# Note: `Intercept` is not an independent variable
which(summary(climate_reg1)$coefficients[, 4] < 0.05)
```

```
## (Intercept)         MEI         CO2      CFC.11      CFC.12
TSI
##           1           2           3           6           7
8
##    Aerosols
##           9
```

The coefficients of `N2O` and `CFC.11` are negative probably because they are correlated with other variables within or beyond the model.

## Correlation

```
cor(climate_train)
```

```
##                 Year         Month           MEI         CO2
CH4
## Year      1.00000000 -0.0279419602 -0.0369876842  0.98274939
0.91565945
## Month    -0.02794196  1.0000000000  0.0008846905 -0.10673246
0.01856866
## MEI      -0.03698768  0.0008846905  1.0000000000 -0.04114717 -
0.03341930
## CO2       0.98274939 -0.1067324607 -0.0411471651  1.00000000
0.87727963
## CH4       0.91565945  0.0185686624 -0.0334193014  0.87727963
1.00000000
## N2O       0.99384523  0.0136315303 -0.0508197755  0.97671982
0.89983864
## CFC.11    0.56910643 -0.0131112236  0.0690004387  0.51405975
0.77990402
## CFC.12    0.89701166  0.0006751102  0.0082855443  0.85268963
0.96361625
## TSI       0.17030201 -0.0346061935 -0.1544919227  0.17742893
0.24552844
## Aerosols -0.34524670  0.0148895406  0.3402377871 -0.35615480 -
0.26780919
## Temp      0.78679714 -0.0998567411  0.1724707512  0.78852921
0.70325502
##                 N2O      CFC.11        CFC.12         TSI
Aerosols
## Year      0.99384523  0.56910643  0.8970116635  0.17030201 -
0.34524670
## Month     0.01363153 -0.01311122  0.0006751102 -0.03460619
0.01488954
## MEI      -0.05081978  0.06900044  0.0082855443 -0.15449192
0.34023779
## CO2       0.97671982  0.51405975  0.8526896272  0.17742893 -
0.35615480
## CH4       0.89983864  0.77990402  0.9636162478  0.24552844 -
0.26780919
## N2O       1.00000000  0.52247732  0.8679307757  0.19975668 -
0.33705457
## CFC.11    0.52247732  1.00000000  0.8689851828  0.27204596 -
0.04392120
## CFC.12    0.86793078  0.86898518  1.0000000000  0.25530281 -
0.22513124
## TSI       0.19975668  0.27204596  0.2553028138  1.00000000
0.05211651
## Aerosols -0.33705457 -0.04392120 -0.2251312440  0.05211651
1.00000000
## Temp      0.77863893  0.40771029  0.6875575483  0.24338269 -
0.38491375
##                Temp
## Year      0.78679714
```

```
## Month    -0.09985674
## MEI       0.17247075
## CO2       0.78852921
## CH4       0.70325502
## N2O       0.77863893
## CFC.11    0.40771029
## CFC.12    0.68755755
## TSI       0.24338269
## Aerosols -0.38491375
## Temp      1.00000000

# `N2O` is highly correlated with
which(cor(climate_train)["N2O", ] > 0.7)

##   Year    CO2    CH4    N2O CFC.12   Temp
##      1      4      5      6      8     11

# Note: `Temp` is the dependent variable

# `CFC.11` is highly correlated with
which(cor(climate_train)["CFC.11", ] > 0.7)

##    CH4 CFC.11 CFC.12
##      5      7      8
```

## Simplify the Model

```
climate_reg2 <- lm(Temp ~ MEI + TSI + Aerosols + N2O,
                   data = climate_train)
summary(climate_reg2)

##
## Call:
## lm(formula = Temp ~ MEI + TSI + Aerosols + N2O, data =
climate_train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.27916 -0.05975 -0.00595  0.05672  0.34195
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.162e+02  2.022e+01  -5.747 2.37e-08 ***
## MEI          6.419e-02  6.652e-03   9.649  < 2e-16 ***
## TSI          7.949e-02  1.487e-02   5.344 1.89e-07 ***
## Aerosols    -1.702e+00  2.180e-01  -7.806 1.19e-13 ***
## N2O          2.532e-02  1.311e-03  19.307  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09547 on 279 degrees of freedom
```

```
## Multiple R-squared:  0.7261, Adjusted R-squared:  0.7222
## F-statistic: 184.9 on 4 and 279 DF,  p-value: < 2.2e-16
```

```
# The sign of `N2O` flips.
c("Previous" = climate_reg1$coefficients["N2O"],
  "Simplified" = climate_reg2$coefficients["N2O"])
```

```
##   Previous.N2O Simplified.N2O
##    -0.01652800     0.02531975
```

```
# The R^2 is lower in the simplified model.
c("Previous R^2" = summary(climate_reg1)$r.squared,
  "Simplified R^2" = summary(climate_reg2)$r.squared)
```

```
##   Previous R^2 Simplified R^2
##      0.7508933      0.7261321
```

```
# Significant independent variables (with p < 0.05)
# Note: `Intercept` is not an independent variable
which(summary(climate_reg2)$coefficients[, 4] < 0.05)
```

```
## (Intercept)         MEI         TSI    Aerosols         N2O
##           1           2           3           4           5
```

A higher proportion of independent variables is significant at 5% (4 out of 4 vs. 6 out of 8).

## Out-of-sample $R^2$

```
climate_predict <- predict(climate_reg2, newdata = climate_test)

SSE <- sum((climate_test$Temp - climate_predict)^2)
SST <- sum((climate_test$Temp - mean(climate_train$Temp))^2)
1 - SSE/SST
```

```
## [1] 0.4967795
```