**Why Model1**

I have chosen model2 because R-Square value is more for model one compared to other model and SSE value of for Model1 Is low when compared to other models. As SSE is one parameter to check quality of a model and here the SSE value is less for model1 I have chosen the model1.

**Computed r-square Result**

0.7630134

As R-square is 0.76 which is nearly 0.8 we can say that the predicted model is accurate .we can go with this model

**The equation is as follows**

Test= -1.231e+02 + MEI (6.367e-02 ) + CO2 (6.906e-03) + N2O( -1.620e-02 ) + CFC.11(-6.410e-03 ) + CFC.12 (3.625e-03) + TSI (9.181e-02) + Aerosols (-1.520e+00 )

**Results and interpretations are below please go through it.**

FOR READING FILE

r<-read.csv('Dataset.csv')

TO SEE STRUCTURE

str(r)

SPLIT OF DATA INTO TRANING

traning<-subset(r, Year < 2006)

traning

SPLIT OF DATA INTO TEST

test<-subset(r,Year>=2006)

test

# Linear Regression (ALL variables)

model1 = lm(Temp ~ MEI+CO2+CH4+N2O+CFC.11+CFC.12+TSI+Aerosols, data=traning)

summary(model1)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.231e+02 2.087e+01 -5.897 1.13e-08 \*\*\*

MEI 6.367e-02 6.685e-03 9.524 < 2e-16 \*\*\*

CO2 6.906e-03 2.395e-03 2.883 0.004262 \*\*

CH4 1.645e-04 5.470e-04 0.301 0.763863

N2O -1.620e-02 9.461e-03 -1.712 0.088083 .

CFC.11 -6.410e-03 1.767e-03 -3.629 0.000342 \*\*\*

CFC.12 3.625e-03 1.104e-03 3.285 0.001159 \*\*

TSI 9.181e-02 1.566e-02 5.861 1.37e-08 \*\*\*

Aerosols -1.520e+00 2.188e-01 -6.949 2.88e-11 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09329 on 263 degrees of freedom

Multiple R-squared: 0.7415, Adjusted R-squared: 0.7337

F-statistic: 94.32 on 8 and 263 DF, p-value: < 2.2e-16

Interpretation:

here I taken all variable and the summary shows that CH4 is less significant and R square is 0.74

# Sum of Squared Errors

model1$residuals

SSE = sum(model1$residuals^2)

SSE

Result

2.288775

SSE Value is a measure of quality of regression and here the value is 2.288

# Linear Regression (7 variables)

model2 = lm(Temp ~ MEI+CO2+N2O+CFC.11+CFC.12+TSI+Aerosols, data=traning)

summary(model2)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.227e+02 2.080e+01 -5.900 1.11e-08 \*\*\*

MEI 6.350e-02 6.650e-03 9.549 < 2e-16 \*\*\*

CO2 6.839e-03 2.381e-03 2.873 0.004402 \*\*

N2O -1.535e-02 9.013e-03 -1.703 0.089806 .

CFC.11 -6.353e-03 1.753e-03 -3.624 0.000348 \*\*\*

CFC.12 3.683e-03 1.085e-03 3.395 0.000793 \*\*\*

TSI 9.155e-02 1.561e-02 5.863 1.35e-08 \*\*\*

Aerosols -1.523e+00 2.182e-01 -6.979 2.40e-11 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09313 on 264 degrees of freedom

Multiple R-squared: 0.7415, Adjusted R-squared: 0.7346

F-statistic: 108.2 on 7 and 264 DF, p-value: < 2.2e-16

Interpretation:

As CH4 is least significant I removied CH4 and created an equation for remaining 7 varibles and Rsquare is 0.74

# Sum of Squared Errors

SSE = sum(model2$residuals^2)

SSE

2.289562

Interpretion

SSE Value is a measure of quality of regression and here the value is 2.29 which is greater than model 1

# Linear Regression (6 variables)

model3 = lm(Temp ~ MEI+CO2+CFC.11+CFC.12+TSI+Aerosols, data=traning)

summary(model3)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.180e+02 2.069e+01 -5.704 3.13e-08 \*\*\*

MEI 6.358e-02 6.674e-03 9.527 < 2e-16 \*\*\*

CO2 4.994e-03 2.128e-03 2.347 0.019639 \*

CFC.11 -4.092e-03 1.149e-03 -3.562 0.000437 \*\*\*

CFC.12 2.237e-03 6.774e-04 3.302 0.001092 \*\*

TSI 8.520e-02 1.522e-02 5.599 5.37e-08 \*\*\*

Aerosols -1.533e+00 2.189e-01 -7.005 2.04e-11 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09346 on 265 degrees of freedom

Multiple R-squared: 0.7386, Adjusted R-squared: 0.7327

F-statistic: 124.8 on 6 and 265 DF, p-value: < 2.2e-16

# Sum of Squared Errors

SSE = sum(model3$residuals^2)

SSE

2.314705

Interpretion

SSE Value is a measure of quality of regression and here the value is 2.314705 which is greater than model2

# Linear Regression (5 variables)

Model4 = lm(Temp ~ MEI+CFC.11+CFC.12+TSI+Aerosols, data=traning)

summary(model4)

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -7.806e+01 2.358e+01 -3.310 0.001062 \*\*

CFC.11 -6.238e-03 7.070e-04 -8.824 < 2e-16 \*\*\*

CFC.12 3.847e-03 2.627e-04 14.641 < 2e-16 \*\*\*

TSI 5.709e-02 1.728e-02 3.305 0.001081 \*\*

Aerosols -8.614e-01 2.365e-01 -3.643 0.000324 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1097 on 267 degrees of freedom

Multiple R-squared: 0.6372, Adjusted R-squared: 0.6317

F-statistic: 117.2 on 4 and 267 DF, p-value: < 2.2e-16

# Sum of Squared Errors

SSE = sum(model4$residuals^2)

SSE

3.213187

INTERPRETION

SSE Value is a measure of quality of regression and here the value is 3.21 which is greater than model3

# Correlations

cor(traning)

cor(traning$CFC.11, traning$CFC.12)

cor(traning$CH4,traning$N2O)

# test set

test<-subset(r,Year>=2006)

test

str(test)

# Make test set predictions

predictTest = predict(model2, newdata=test)

predictTest

plot(predictTest)

# Compute R-squared

SSE = sum((test$Temp - predictTest)^2)

SST = sum((test$Temp - mean(traning$Temp))^2)

1 - SSE/SST

Result

0.7630134

As R-square is 0.76 which is nearly 0.8 we can say that the predicted modelis accurate .we can go with this model