STA137 Project I

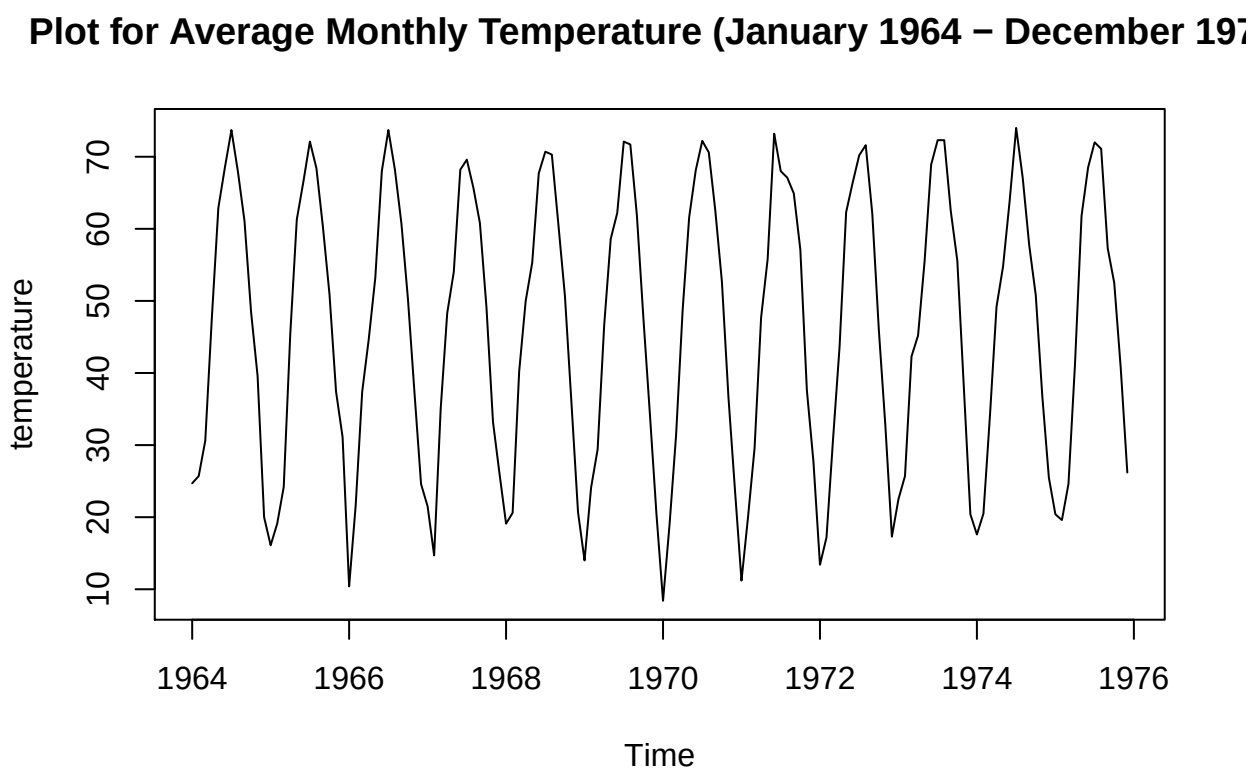
Junyao Lu SID:915938649

28/10/2020

```
data = read.csv("temperature.csv")
df = ts(data, frequency = 12, start = c(1964, 1))
```

(a)

```
plot(df, main = "Plot for Average Monthly Temperature (January 1964 - December 1975)")
```



From the plot above, there is not trend observed. However, the plot shows a seasonal pattern in the data, which roughly repeats every year.

(b)

```
M = factor(cycle(df))
reg1 = lm(df ~ 0 + M, na.action=NULL)
summary(reg1)
```

```
##
## Call:
## lm(formula = df ~ 0 + M, na.action = NULL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2750 -2.2479  0.1125  1.8896  9.8250
##
```

```
## Coefficients:
##     Estimate Std. Error t value Pr(>|t|)
## M1    16.608      0.987   16.83   <2e-16 ***
## M2    20.650      0.987   20.92   <2e-16 ***
## M3    32.475      0.987   32.90   <2e-16 ***
## M4    46.525      0.987   47.14   <2e-16 ***
## M5    58.092      0.987   58.86   <2e-16 ***
## M6    67.500      0.987   68.39   <2e-16 ***
## M7    71.717      0.987   72.66   <2e-16 ***
## M8    69.333      0.987   70.25   <2e-16 ***
## M9    61.025      0.987   61.83   <2e-16 ***
## M10   50.975      0.987   51.65   <2e-16 ***
## M11   36.650      0.987   37.13   <2e-16 ***
## M12   23.642      0.987   23.95   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.419 on 132 degrees of freedom
## Multiple R-squared:  0.9957, Adjusted R-squared:  0.9953
## F-statistic:  2569 on 12 and 132 DF,  p-value: < 2.2e-16
```

(i)

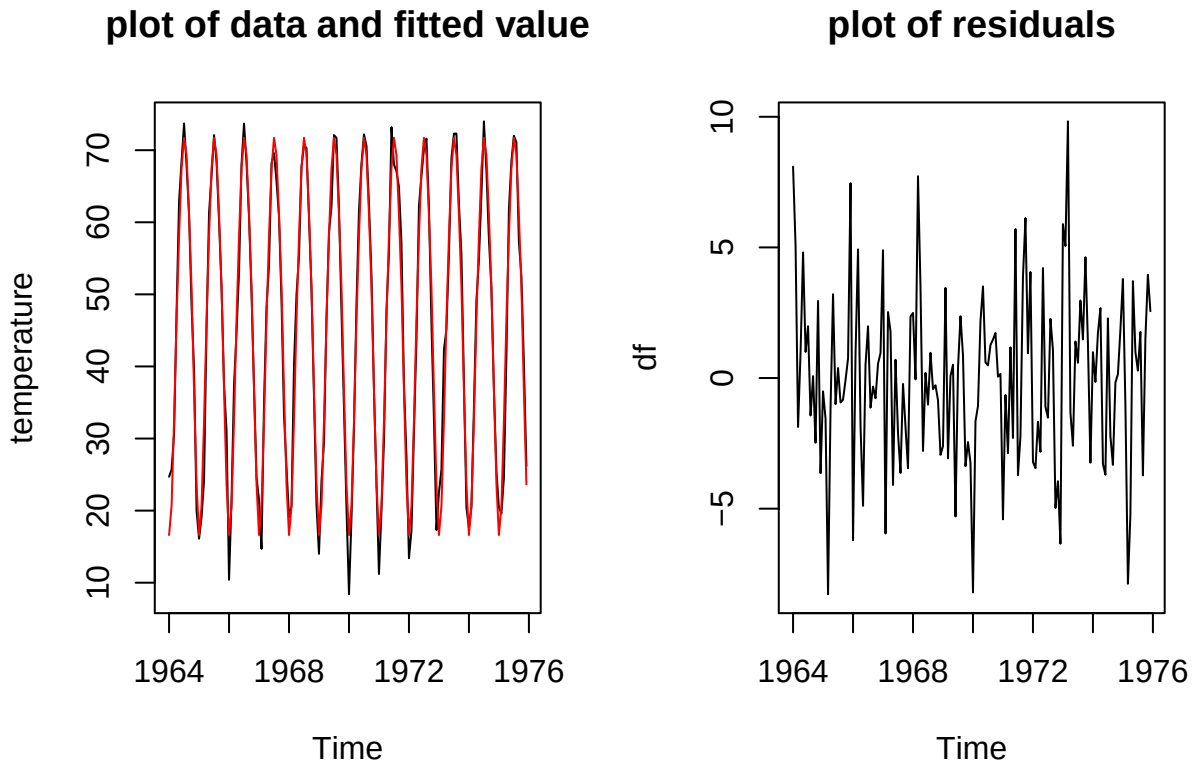
```
diff = vector()
for (i in 1:11){
  diff = c(diff, reg1$coefficients[[i+1]] - reg1$coefficients[[i]])
}
diff
```

```
##  [1]   4.041667  11.825000  14.050000  11.566667   9.408333   4.216667
##  [7]  -2.383333  -8.308333 -10.050000 -14.325000 -13.008333
```

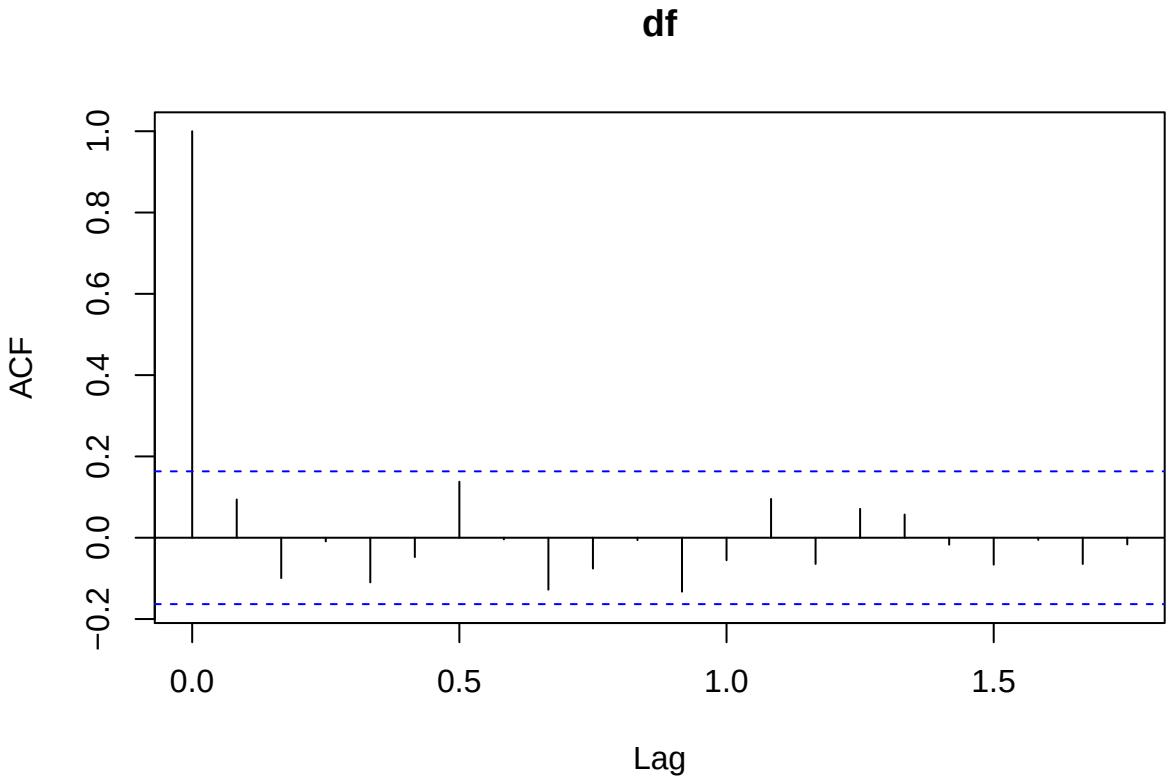
The average temperature increases by 4.041667 from January to February; The average temperature increases by 11.825000 from February to March; The average temperature increases by 14.050000 from March to April; The average temperature increases by 11.566667 from April to May; The average temperature increases by 9.408333 from May to June; The average temperature increases by 4.216667 from June to July; The average temperature decreases by 2.383333 from July to August; The average temperature decreases by 8.308333 from August to September; The average temperature decreases by 10.050000 from September to October; The average temperature decreases by 14.325000 from October to November; The average temperature decreases by 13.008333 from November to December.

(ii)

```
par(mfrow=c(1,2))
plot(df, main="plot of data and fitted value")
lines(fitted(reg1), col="red")
plot(df-fitted(reg1), main="plot of residuals")
```




```
par(mfrow=c(1,1))
acf(df-fitted(reg1))
```



From the plots above, it appears that the model fits the data well and the residuals look white. The plot of

data and the plot of fitted values match with each other very well and the residuals seem to be random and ACF drops immediately from 1 to 0.

(c)

```
sincos = vector()
t = 1:144
for(k in 1:6){
  sincos = cbind(sincos, cos(2*pi*k*t*1/12), sin(2*pi*k*t*1/12))
}
sincos = as.data.frame(sincos)
names(sincos) = c("cos1", "sin1", "cos2", "sin2",
                  "cos3", "sin3", "cos4", "sin4", "cos5", "sin5",
                  "cos6", "sin6")

reg2 = lm(df~sincos$cos1+sincos$sin1+sincos$cos2+sincos$sin2
          +sincos$cos3+sincos$sin3+sincos$cos4+sincos$sin4
          +sincos$cos5+sincos$sin5+sincos$cos6+sincos$sin6, na.action=NULL)
summary(reg2)
```

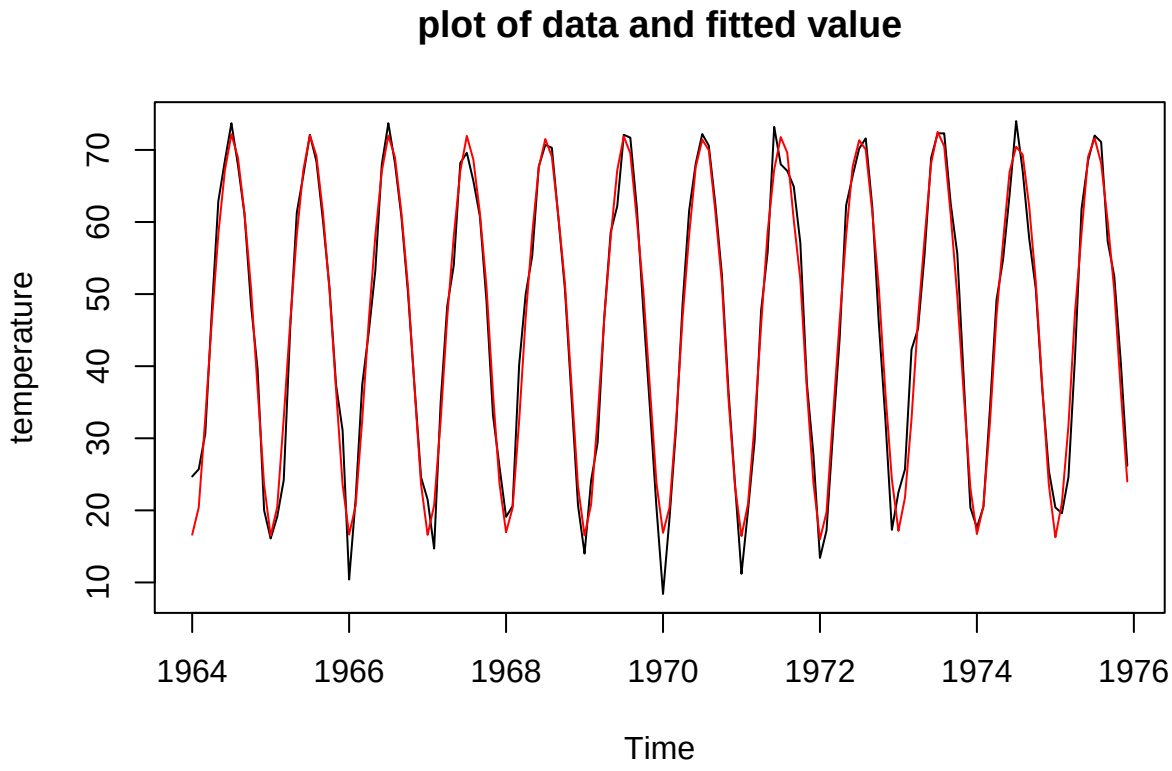
```
##
## Call:
## lm(formula = df ~ sincos$cos1 + sincos$sin1 + sincos$cos2 + sincos$sin2 +
##     sincos$cos3 + sincos$sin3 + sincos$cos4 + sincos$sin4 + sincos$cos5 +
##     sincos$sin5 + sincos$cos6 + sincos$sin6, na.action = NULL)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -8.571 -2.319  0.112  1.928  9.661
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.624e+01  2.834e-01 163.182  < 2e-16 ***
## sincos$cos1 -2.207e+01  4.004e-01 -55.121  < 2e-16 ***
## sincos$sin1 -1.524e+01  4.002e-01 -38.094  < 2e-16 ***
## sincos$cos2 -7.251e-01  4.007e-01  -1.810   0.0726 .
## sincos$sin2 -2.026e+00  4.002e-01  -5.062 1.38e-06 ***
## sincos$cos3  6.043e-02  4.001e-01   0.151   0.8802
## sincos$sin3 -9.155e-01  4.018e-01  -2.278   0.0243 *
## sincos$cos4 -8.006e-02  4.004e-01  -0.200   0.8418
## sincos$sin4  1.667e-01  4.002e-01   0.417   0.6777
## sincos$cos5  1.868e-01  4.078e-01   0.458   0.6477
## sincos$sin5  6.016e-02  4.010e-01   0.150   0.8810
## sincos$cos6 -1.065e-01  3.272e-01  -0.325   0.7455
## sincos$sin6 -2.828e+13  1.673e+13  -1.690   0.0933 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.395 on 131 degrees of freedom
## Multiple R-squared:  0.9718, Adjusted R-squared:  0.9693
## F-statistic: 376.8 on 12 and 131 DF,  p-value: < 2.2e-16
```

(i)

The significant frequency is f = 1/12. It takes 12 months for the time series to repeat itself.

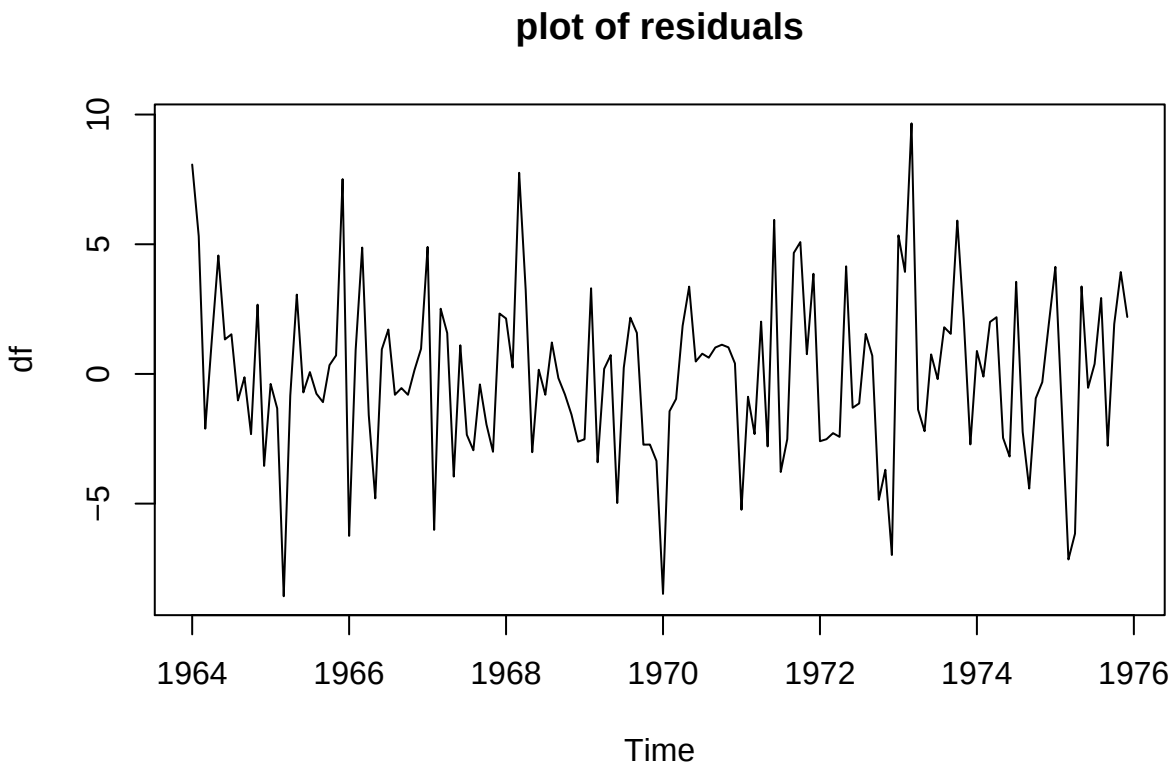
(ii)

```
par(mfrow=c(1,1))
plot(df, main="plot of data and fitted value")
lines(fitted(reg2), col="red")
```

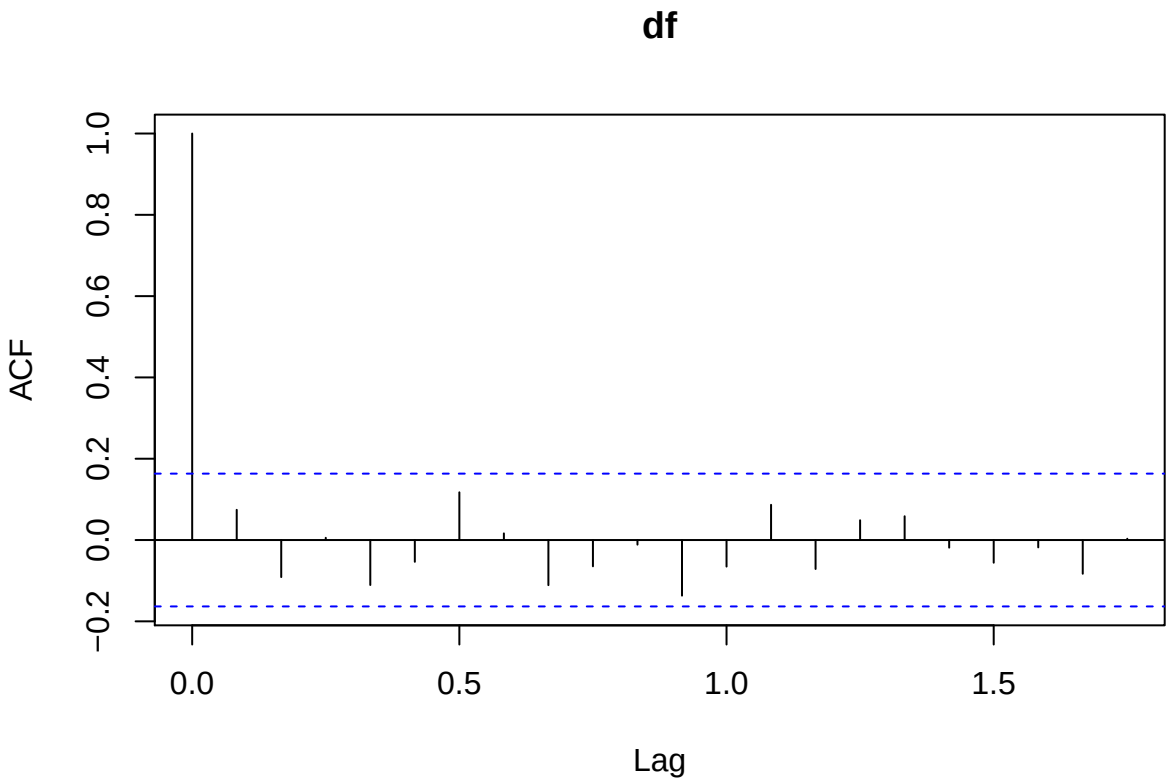


(iii)

```
plot(df-fitted(reg2), main="plot of residuals")
```




```
acf(df-fitted(reg2))
```



From the plots above, it appears that the model fits the data well and the residuals look white. However, it appears that the residuals are not a sample from a Normal distribution, because there is not any trend shown in the plot of residuals.

Code Appendix

```
data = read.csv("temperature.csv")
df = ts(data, frequency = 12, start = c(1964, 1))
plot(df, main = "Plot for Average Monthly Temperature (January 1964 - December 1975)")
M = factor(cycle(df))
reg1 = lm(df ~ 0 + M, na.action=NULL)
summary(reg1)
diff = vector()
for (i in 1:11){
  diff = c(diff, reg1$coefficients[[i+1]] - reg1$coefficients[[i]])
}
diff
par(mfrow=c(1,2))
plot(df, main="plot of data and fitted value")
lines(fitted(reg1), col="red")
plot(df-fitted(reg1), main="plot of residuals")
par(mfrow=c(1,1))
acf(df-fitted(reg1))
sincos = vector()
t = 1:144
for(k in 1:6){
  sincos = cbind(sincos, cos(2*pi*k*t*1/12), sin(2*pi*k*t*1/12))
}
sincos = as.data.frame(sincos)
names(sincos) = c("cos1", "sin1", "cos2", "sin2",
                  "cos3", "sin3", "cos4", "sin4", "cos5", "sin5",
                  "cos6", "sin6")

reg2 = lm(df~sincos$cos1+sincos$sin1+sincos$cos2+sincos$sin2
          +sincos$cos3+sincos$sin3+sincos$cos4+sincos$sin4
          +sincos$cos5+sincos$sin5+sincos$cos6+sincos$sin6, na.action=NULL)
summary(reg2)
par(mfrow=c(1,1))
plot(df, main="plot of data and fitted value")
lines(fitted(reg2), col="red")
plot(df-fitted(reg2), main="plot of residuals")
acf(df-fitted(reg2))
```