Temperature Trend in Chennai over 1990 to 2022

Harsh Mittal

Read the data

```
data <- read.csv("C:/Users/harsh.hm.mittal/OneDrive/Desktop/Predictive Analytics - Regression and Classi
str(data)
```

```
## 'data.frame':    11894 obs. of  5 variables:
##  $ time: Factor w/ 11894 levels "01-01-1990","01-01-1991",..: 1 392 783 1174 1565 1956 2347 2738 3129
##  $ tavg: num  25.2 24.9 25.6 25.7 25.5 24.7 25.4 25.6 24.8 24.7 ...
##  $ tmin: num  22.8 21.7 21.4 NA 20.7 NA 23.3 22 21.7 20.7 ...
##  $ tmax: num  28.4 29.1 29.8 28.7 28.4 26.1 27 28 28.5 29 ...
##  $ prcp: num  0.5 0 0 0 0 0.5 18 0.5 0 0 ...
```

```
data$time=as.Date(data$time,format = "%d-%m-%Y")
str(data)
```

```
## 'data.frame':    11894 obs. of  5 variables:
##  $ time: Date, format: "1990-01-01" "1990-01-02" ...
##  $ tavg: num  25.2 24.9 25.6 25.7 25.5 24.7 25.4 25.6 24.8 24.7 ...
##  $ tmin: num  22.8 21.7 21.4 NA 20.7 NA 23.3 22 21.7 20.7 ...
##  $ tmax: num  28.4 29.1 29.8 28.7 28.4 26.1 27 28 28.5 29 ...
##  $ prcp: num  0.5 0 0 0 0 0.5 18 0.5 0 0 ...
```

```
n = nrow(data)
data$tms = 1:n
data$tms = data$tms - mean(data$tms)
```

Split the data into train and test

```
data_train = subset(data,time<=as.Date("2015-12-31"))
tail(data_train)
```

```
##            time tavg tmin tmax prcp    tms
## 9491 2015-12-26 25.8 21.3   31   NA 3543.5
## 9492 2015-12-27 26.7 23.0   31    0 3544.5
## 9493 2015-12-28 26.3 22.8   31   NA 3545.5
## 9494 2015-12-29 26.1 21.3   31   NA 3546.5
## 9495 2015-12-30 26.1 20.3   31   NA 3547.5
## 9496 2015-12-31 26.3 20.8   31   NA 3548.5
```

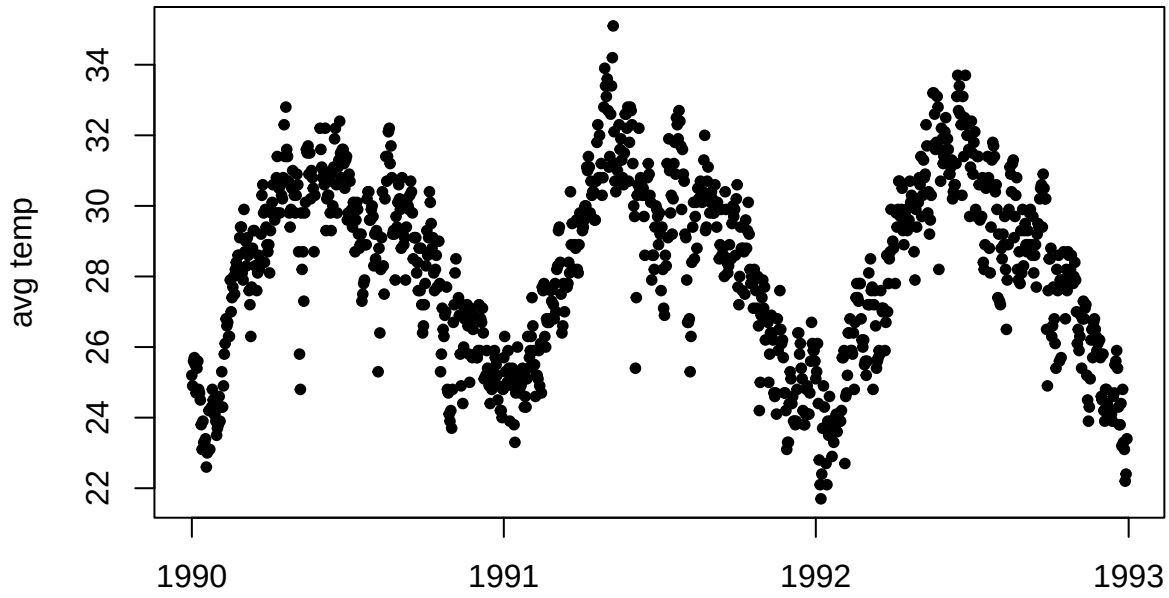
```
data_test = subset(data,time>as.Date("2015-12-31"))
head(data_test)
```

```
##            time tavg tmin tmax prcp    tms
## 9497 2016-01-01 26.3 21.2 30.2   NA 3549.5
## 9498 2016-01-02 25.8 20.6 31.0   NA 3550.5
## 9499 2016-01-03 25.9 19.6 32.0   NA 3551.5
```

```
## 9500 2016-01-04 26.6 20.5 32.0   NA 3552.5
## 9501 2016-01-05 26.1 20.5 31.3   NA 3553.5
## 9502 2016-01-06 26.4 20.5 31.0   NA 3554.5
```

visualisation

```
plot(data_train$time[1:(3*365)],data_train$tavg[1:(3*365)],pch=20,
     xlab='',ylab='avg temp')
```



Fit model 1

$$tavg = \beta_0 + \beta_1 t + \beta_2 \sin(\omega t) + \beta_3 \cos(\omega t) + \varepsilon,$$

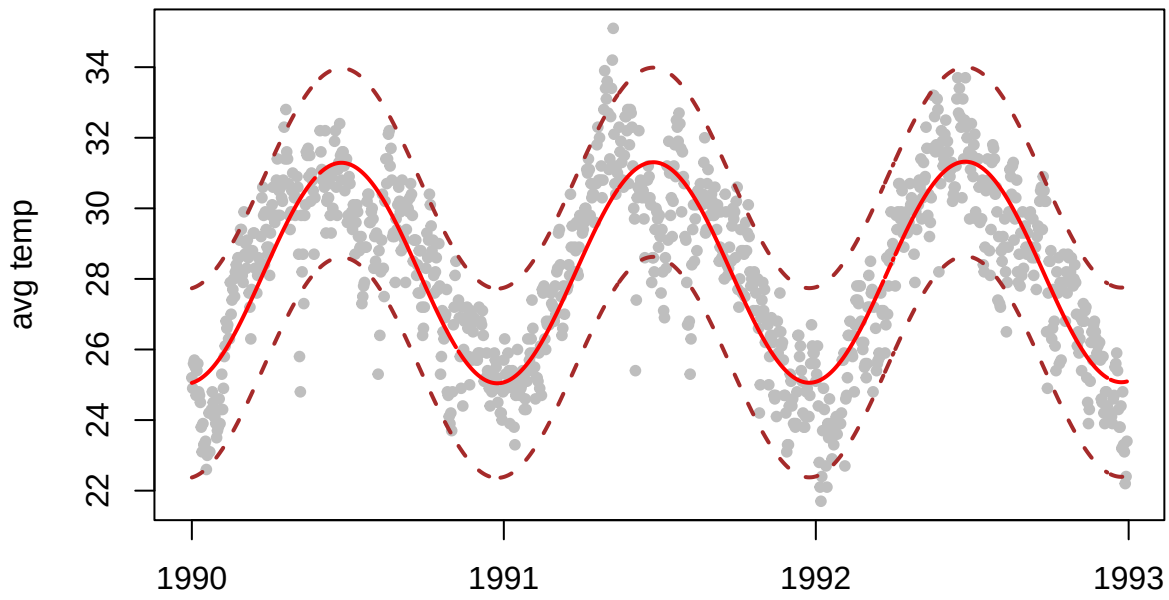
where $\omega = \frac{2\pi}{P}$, $P = 365$.

```
## modl 1

omega = 2*pi/365
mod1 = lm(tavg ~ tms + sin(omega*tms) + cos(omega*tms), data = data_train)
sum = summary(mod1)
sum
```

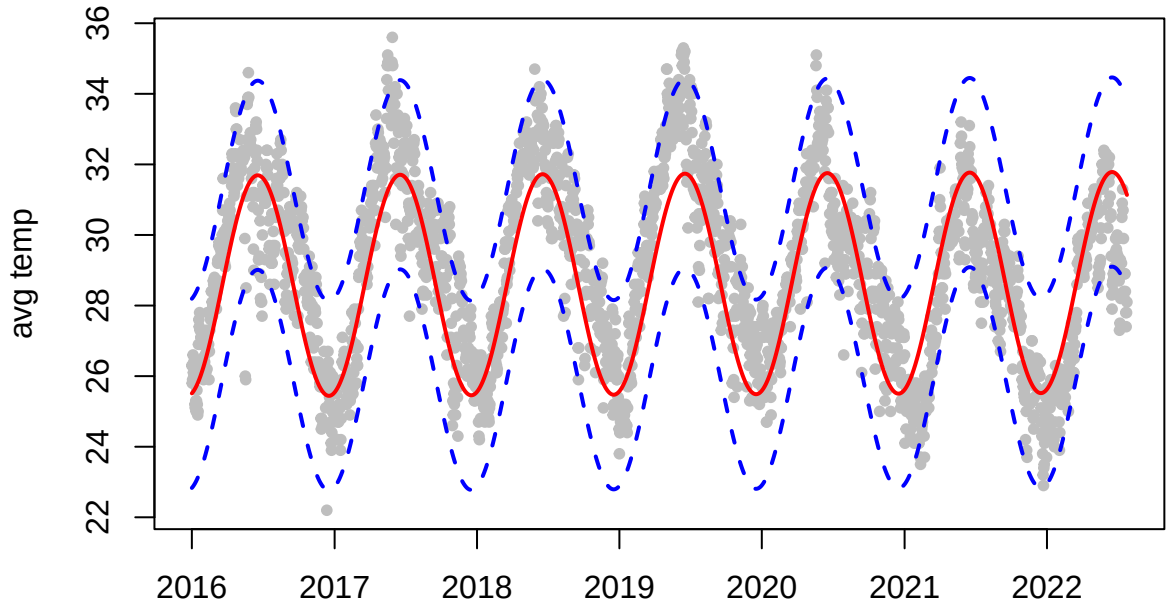
```
##
## Call:
## lm(formula = tavg ~ tms + sin(omega * tms) + cos(omega * tms),
##     data = data_train)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.9459 -0.8738  0.0368  0.9325  5.5168
##
## Coefficients:
##                   Estimate Std. Error  t value Pr(>|t|)
## (Intercept)      2.841e+01  1.533e-02 1853.481  < 2e-16 ***
## tms              4.216e-05  5.135e-06    8.211 2.49e-16 ***
## sin(omega * tms) 2.889e+00  1.987e-02  145.404  < 2e-16 ***
## cos(omega * tms) 1.199e+00  1.990e-02   60.271  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.368 on 9465 degrees of freedom
##   (27 observations deleted due to missingness)
## Multiple R-squared:  0.7239, Adjusted R-squared:  0.7238
## F-statistic:  8273 on 3 and 9465 DF,  p-value: < 2.2e-16
```

```
sigma=sum$sigma

data_train$fitted.values=NA #creaing a new col with all NAs
data_train[rownames(mod1$model),'fitted.values']=mod1$fitted.values #filling that col with fitted value

plot(data_train$time[1:(3*365)], data_train$tavg[1:(3*365)]
     ,pch=20,col='grey'
       ,xlab='',ylab='avg temp')

lines(data_train$time[1:(3*365)], data_train$fitted.values[1:(3*365)]
      ,col='red',lwd=2)

lines(data_train$time[1:(3*365)]
      ,data_train$fitted.values[1:(3*365)]-1.96*sigma
      ,col='brown',lwd=2,lty=2)

lines(data_train$time[1:(3*365)]
      ,data_train$fitted.values[1:(3*365)]+1.96*sigma
      ,col='brown',lwd=2,lty=2)
```




```
#### do prediction in test data


data_test$pred = predict(mod1,newdata = data_test)


plot(data_test$time, data_test$tavg
     ,pch=20,col='grey'
       ,xlab='',ylab='avg temp')

lines(data_test$time
      ,data_test$pred
      ,col='red',lwd=2)

lines(data_test$time
      ,data_test$pred-1.96*sigma
      ,col='blue',lwd=2,lty=2)
lines(data_test$time
      ,data_test$pred+1.96*sigma
      ,col='blue',lwd=2,lty=2)
```




```
#sin and cos are engineered features
```