Rainfall TS Analysis

Team 9

2023-12-01

```r
#Data Cleaning
library(readr)
library(tidyverse)
library(dplyr)
data = read_csv('Perth_Rainfall.csv')
data$Month <- format(as.Date(paste(data$Month, sep = "","-01"), format = "%y-%b-%d"),"%Y-%m")
data_ts = ts(data$Rainfall)
data = data.frame(ts(data))
plot(x = data$Month, y = data$Rainfall, xlab = "Month", type = "o",
     ylab = "Rainfall (mm)", main = "Monthly Average Rainfall in Perth")


## Box Jenkins Method
library(TSA)
#Sample ACF
acf(data_ts, lag = length(data_ts), main="Perth Monthly Average
Rainfall\n(ACF)")

#Sample ACF (Two Years)
acf(data_ts, lag = 24, main="Perth Monthly Average Rainfall\n(24 Months)",xlab="Lag (Months)")

#Sample PACF
pacf(data_ts, lag = 24, main= "Perth Monthly Average Rainfall\n(PACF)", xlab="Lag (Months)")

#Sample EACF
eacf(data_ts) ## Maybe a ARMA(3,2), possibly (2,1)


#Dynamics Method
library(TSA)
library(forecast)

#ARMA(1,0)
fit10 <- arima(data_ts, order = c(1,0,0))
fit10
ggtsdisplay(fit10$residuals, lag.max=107, main = "ARMA(1,0)")
checkresiduals(fit10)

#ARMA(2,1)
fit21 <- arima(data_ts, order = c(2,0,1))
fit21
ggtsdisplay(fit21$residuals, lag.max=107, main = "ARMA(2,1)")
checkresiduals(fit21)
```

```r
#ARMA(3,2)
fit32 <- arima(data_ts, order = c(3,0,2))
fit32
ggtsdisplay(fit32$residuals, lag.max=107, main = "ARMA(3,2)")
checkresiduals(fit32)


#ARMA(2,1)x(1,0)
fit2110 <- arima(data_ts, order = c(2,0,1), seasonal = list(order = c(1,0,0), period = 12))
fit2110
ggtsdisplay(fit2110$residuals, lag.max=107, main = "ARMA(2,1)x(1,0)12")
checkresiduals(fit2110)

#ARMA(2,1)x(2,1)
fit2121 <- arima(data_ts, order = c(2,0,1), seasonal = list(order = c(2,0,1), period = 12))
fit2121
ggtsdisplay(fit2121$residuals, lag.max=107, main = "ARMA(2,1)x(2,1)12")
checkresiduals(fit2121)

#Forecasting
fore2110 <- forecast(data_ts,model = fit2110, h = 12)
plot(fore2110, xlab = "Month", ylab = "Precipitation (mm)",main = "12 Month
Precipitation Forecast Using ARMA(2,1)x(1,0)12")

#Validating
library(readr)

new<- read_csv("IDCJAC0009_009021_1800_Data.csv") #daily dataset
new = new[, c(3,4,5,6)]
new = new[new$Year > "2007",]
new$Date = paste(new$Year, new$Month, new$Day, sep = "-")

new$Date <- as.Date(new$Date)
new$Month <- format(new$Date, "%Y-%m")
new <- new[new$Date >= "2008-07-01" & new$Date <= "2018-06-25",]
# original data is ended at 2017-06-25 so one more year
original = new[new$Date >= "2008-07-01" & new$Date <= "2017-06-25",]
nrow(original)

new = new[,c(2,4)]
original = original[,c(2,4)]
colnames(new) = c("Month", "Rainfall")
colnames(original) = c("Month", "Rainfall")
# new %>% is.na() %>% sum() no NA


new <- aggregate(Rainfall ~ Month, data = new, FUN = mean, na.rm = TRUE)
nrow(new)
original = aggregate(Rainfall ~ Month, data = original, FUN = mean, na.rm = TRUE)
nrow(original)

new_ts = ts(new$Rainfall)
original_ts = ts(original$Rainfall)
```

```r
new = data.frame(ts(new))
original = data.frame(ts(original))

#plot(x = new$Month, y = new$Rainfall, xlab = "Month", type = "o",
#      ylab = "Rainfall (mm)", main = "Actual Monthly Average Rainfall in Perth #(108 + 12 Month)")
#points(x = original$Month, y = original$Rainfall, col = "blue", pch = 16, type = "l")
##points(x = data$Month, y = data$Rainfall, col = "red", pch = 16, type = "l")
#abline(v = 108, col = "blue", lty = 2)
#legend("topright", legend = c( "Extended Period"),
#       col = c( "blue"), pch = 16, lty =2, cex = 0.8)
```