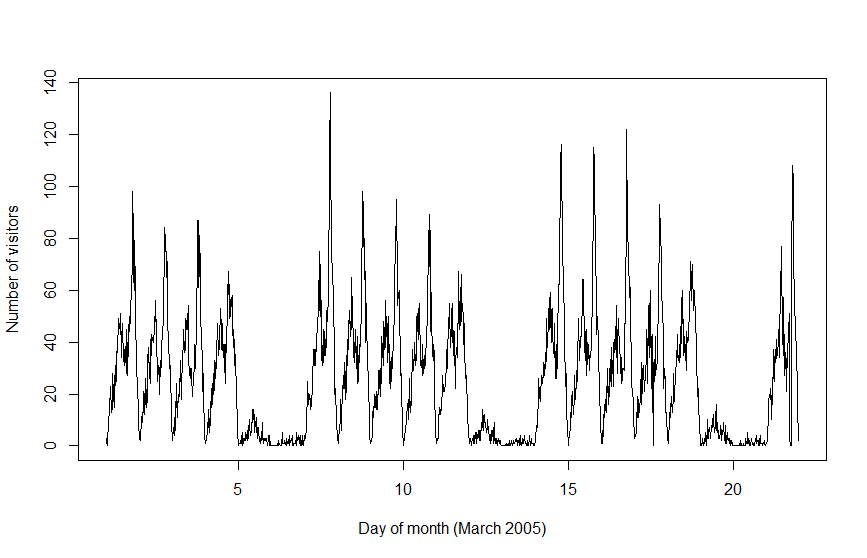
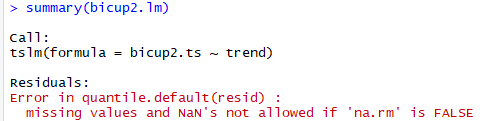
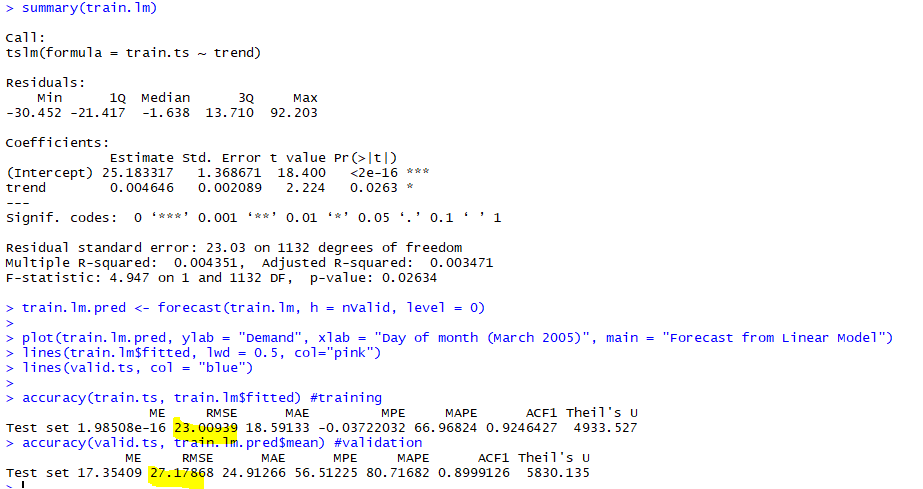
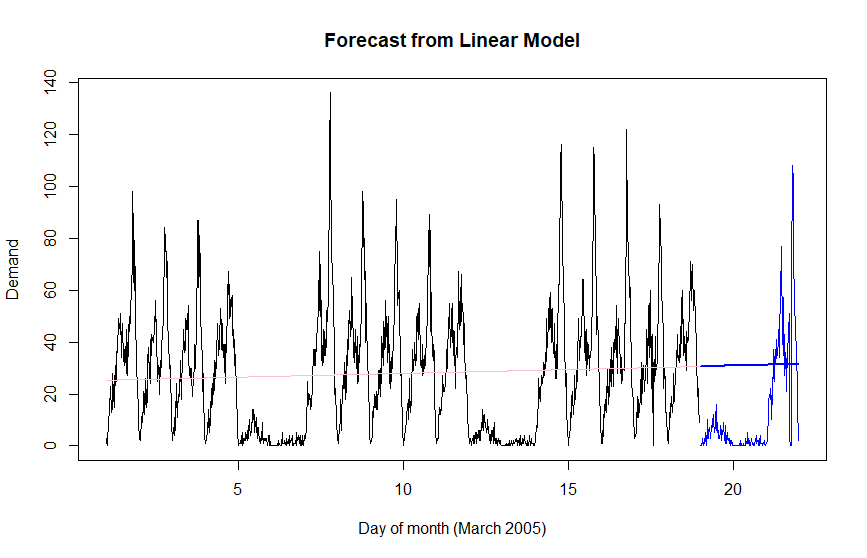
**יישומי R בכריית נתונים, תרגיל 4 – קובץ פלט**

1. פלט הגרף המתקבל מהרצת נתוני המידע ההיסטורי בלבד. לתאריכים 1.3.05-21.05.2005 כולל.
2. פירוט רכיבי הסדרה העתית, לפי גרף הפלט בסעיף 2.
3. **Level-** הסדרה מתחילה מערך 1 שהוא מאוד קרוב ל-0, יחד עם זאת לא מתחילה מ-0 מוחלט.
4. **Trend-** לינארי או אקספוננציאלי? בחינה ויזואלית **לא** מעידה על מגמה אקספוננציאלית ו/או לינארית. נבחן את המגמה על ידי התאמת ה-trend והשוואת ערכי ה-RMSE. פירוט בסעיף 4.
5. **Seasonality-** לסדרה עונתיות מורכבת, אנחנו רואים מחזוריות פחות או יותר דומה בכל אחד מימות השבוע, ואילו מחזוריות נפרדת בימים 6-7 של כל שבוע (סופי שבוע?). כלומר- פעם אחת יש מחזוריות בכל יום ופעם שניה יש מחזוריות בכל שבוע.
6. **Noise**- תצפית 428, מספר חריג של מבקרים בתאריך 7.3.05 בשעה 18:45- 136 מבקרים. נוריד תצפית זו בעזרת יצירת סדרה עתית אשר אינה מכילה אותה.
7. מהתאמת מודל רגרסיה ל-trend לינארי ואקספוננציאלי התקבלו אותם ערכי השגיאות בשני המודלים. הרצת summary נתנה פלט שגיאה, קרי לא ניתן להצביע על trend מובהק, בשונה ממסד ה-cocacola.

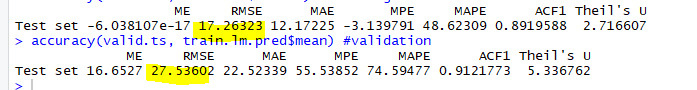


תחזית ביקוש בשלושת הימים החסרים, לפי טבלת ה-summary, רואים ערכי trend ו-intercept.

ביצועי המודל (לינארי), לפי RMSE.



1. התאמת עונתיות **פשוטה** למודל, נותנת את ערכי הפלט. התאריכים אותם אנחנו חוזים 22-24.5.05 לפי התחזית, הם ימי חול (בהם ההתנהגות "רגילה") ולכן עונתיות פשוטה יכולה לתת מענה לצורך המבוקש.



קובץ R:



setwd("C:/Users/dom/Desktop/r files")

bicup.df<-read.csv("bicup2006.csv")

head(bicup.df)

library(forecast)

library(lubridate)

bicup.ts <- ts(bicup.df$DEMAND,

start = c(1, 1), end = c(21, 63), freq = 63)

View(bicup.ts)

plot(bicup.ts, xlab = "Day of month (March 2005)", ylab = "Number of visitors")

#w/o noise

bicup2.ts <- ts(bicup.df$DEMAND[-c(428:428)],

start = c(1, 1), end = c(21, 63), freq = 63)

View(bicup2.ts)

plot(bicup2.ts, xlab = "Day of month (March 2005)", ylab = "Number of visitors")

#fit trend linear

bicup2.lm <- tslm(bicup2.ts ~ trend)

lines(bicup2.lm$fitted, lwd = 0.5, col = "green")

summary(bicup2.lm)

accuracy(bicup2.lm)

#fit trend exponential

bicup2.lm <- tslm(bicup2.ts ~ trend, lambda = 0)

lines(bicup2.lm$fitted, lwd = 0.5, col = "black")

summary(bicup2.lm)

accuracy(bicup2.lm)

nValid <- 189

nTrain <- length(bicup2.ts) - nValid

train.ts <- window(bicup2.ts, start = c(1, 1), end = c(1, nTrain))

valid.ts <- window(bicup2.ts, start = c(1, nTrain+1), end = c(1, nTrain + nValid))

train.lm <- tslm(train.ts ~ trend)

summary(train.lm)

train.lm.pred <- forecast(train.lm, h = nValid, level = 0)

plot(train.lm.pred, ylab = "Demand", xlab = "Day of month (March 2005)", main = "Forecast from Linear Model")

lines(train.lm$fitted, lwd = 0.5, col="pink")

lines(valid.ts, col = "blue")

accuracy(train.ts, train.lm$fitted) #training

accuracy(valid.ts, train.lm.pred$mean) #validation

#fit seasonality

bicup2.lm <- tslm(bicup2.ts ~ trend+season)

lines(bicup2.lm$fitted, lwd = 0.5, col = "red")

nValid <- 189

nTrain <- length(bicup2.ts) - nValid

train.ts <- window(bicup2.ts, start = c(1, 1), end = c(1, nTrain))

valid.ts <- window(bicup2.ts, start = c(1, nTrain+1), end = c(1, nTrain + nValid))

train.lm <- tslm(train.ts ~ trend+season)

summary(train.lm)

train.lm.pred <- forecast(train.lm, h = nValid, level = 0)

plot(train.lm.pred, ylab = "Demand", xlab = "Day of month (March 2005)", main = "Forecast from Linear Model")

lines(train.lm$fitted, lwd = 0.5, col="pink")

lines(valid.ts, col = "blue")

accuracy(train.ts, train.lm$fitted) #training

accuracy(valid.ts, train.lm.pred$mean) #validation