

LABORATION 3

732G52 - HT2024

Intro

Innehåll:

- ARIMA-modeller

Denna laboration är till för er övning på kursmaterialet. Det finns ingen obligatorisk inlämning för den. Uppgifterna utgår från R och Rstudio. Om du behöver repetition i R:

- Titta på kurshemsidan för 732G33: [här](#).
- Cheat sheets [här](#)

Material:

- Introduction to Time Series Analysis and Forecasting (TSAF): se R-kod i boken
- Forecasting: Principles and Practice (FPP), tredje upplagan: [länk](#)
- Repo för kursen: [länk](#)
- Om man behöver repetition på vanlig linjär regression, se kursen Linjära modeller 1, titta tex [här](#): [länk](#)

Se sektionen "The `lm()` function" på sidan 4 för detaljer kring funktionen `lm()`.

Uppgifter

1. Simulera data från ARMA-modeller: Gå igenom koden som finns här - [länk](#)
Tanken är att ni ska bekanta er mer ARMA-familjen genom att simulera tidserier med kända parametervärden och sedan studera tidseriegrafnen, den teoretiska autokorrelationen och den teoretiska partiella autocorrelationen.
2. FPP3 kap 9: ARIMA models
 - (a) Gå igenom och återskapa koden i 9.1 (ej unit root test)
 - (b) Gå igenom och återskapa koden i 9.5
 - (c) Gå igenom och återskapa koden i 9.7 (ej Plotting the characteristic roots)
 - (d) Gå igenom och återskapa koden i 9.9
 - (e) Gå igenom och återskapa koden i 9.10
 - Här finns kod relaterad till kap 9: [länk1](#), [länk2](#)
3. FPP3 9.11 Exercies: Gör uppgifterna:
 - 6, 7, 9, 11, 15
4. Ladda in datamaterialet `hare` eller `color` som finns i paketet `TSA`. Ta reda på vad data handlar om och gör en lämplig tidserieplot. Vad har data för egenskaper? Avgör om tidserien är stationär eller ej. Om inte genomför lämpliga differanser. Studera SAC och SPAC och välj en lämplig ARIMA-modell för tidserien. Anpassa den valda modellen. Plotta anpassade värden tillsammans med data. Studera residualerna. Gör prognos 5 tidssteg framåt.
5. Ni ska studera antalet nya lastbilsregistreringar, se filen "`NyLastbilregistreringar.csv`" som finns här: [länk](#)
 - (a) Plotta först data. Eftersom det är månadsdata med tydlig säsongsvariation så differentierar du serien för månad och sparar den i en kolumn som du kallar `diff12`.
 - (b) Plotta nu `diff12` samt skatta `acf`, dvs beräkna SAC. Dessa båda grafer tyder på att du behöver differentiera `diff12` en gång för trend. Gör det och kalla denna diffade serie `diff12_1`.
 - (c) Plotta denna serie samt beräkna SAC. Visst ser den stationär ut! Beräkna även SPAC på `diff12_1` och bestäm vilken ARIMA modell som är lämplig.
 - (d) Skatta den den valda modellen och skapa lämpliga residualplottar.

- (e) Är du nöjd med modellen? (Det är svårt att få riktigt bra residualer). Om inte, testa några olika modeller. När du är nöjd, ta fram prognoser för 12 månader fram i tiden. Korrigera nu dessa prognoser så att du får prognoser för antal nylastbilsregistreringar (alltså den odiffade serien) . Kommentera.
6. Turist-data: utgå från filen "turisterTurkiet.csv". Beskrivningen av data finns i "tu_foreign_visitors_93-04.txt". Filerna ligger här: [länk](#)
- (a) Logaritmera serien och plotta den.
 - (b) Målet är nu att anpassa en lämplig SARIMA-modell. Studera SAP och SPAC, och testa olika diffar om ni anser att det behövs. Använd kunskapen från dessa undersökningar för att välja en lämplig SARIMA-modell.
 - (c) Anpassa modellen och studera residualerna. Är ni nöjd? Om inte försök att justera modellen. När ni är nöjda, gör prognos 12 månader framåt.
7. Det finns flera alternativ till paketet **fable** om man arbetar med ARIMA/SARIMA-modeller i R. Några exempel: I base R så finns funktionerna **ar()** och **arima()**. I paket **forecast** finns det funktioner för ARIMA-modellering. Den andra upplagan av boken FPP använder paketet **forecast**.

The `lm()` function

The function `lm()` will be important during the course. Check out the documentation with `?lm()`.

Note that R is an object oriented language, and the `lm()` returns objects with class “`lm`”, with has the form of a list, so you can easily fetch different part of the object when needed. These objects has several useful generic functions connected to it:

- `coef()`: Gives the regression coefficients
- `residuals()`: calculates the residuals of the model
- `fitted()`: Gives the fitted values of the model
- `summary()`: give detail summary and inference. It will return a object of class “`summary.lm`”. `coef()` will work on this object.
- `anova()`: Calculates anova table for the model
- `predict()`: make predictions with the model for (new) data
- `plot()`: output diagnostics plots for the model

In general, for documentation for these methods run commands of the type `?summary.lm()` in the terminal. Another useful function is to use `str()` on the `lm`-object, to get detailed information about it.