

LABORATION 4

732G52 - HT2025

Intro

Innehåll:

- Utvärdering av prognoser
- Tidsseriegression med ARIMA
- Några fall av vanliga problem i tidserieanalys: flera säsonger, tidsserier med en mindre tidsskala än månader, outliers och saknade värden.

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:

- Time Series for Data Science: se R-kod, dataset mm: länk
- 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

Uppgifter

1. Om man inte har gjort det tidigare: Gå igenom text och kod i 5.8 och 5.10 i FPP3 kap 5: The forecaster's toolbox. Här finns länkar till kod, men se till att läsa i boken också: länk1, länk2,
2. FPP3: 5.11 Exercises, gör uppgift: 7
3. Gå igenom koden här: länk. Vad händer i koden? Vilka modeller skattas? Vilken modell verkar bäst för datasetet?
4. Ni ska studera antalet nya lastbilsregistreringar, se filen "NyLastbilregistreringar.csv". som finns här: länk. Målet är att skatta olika modeller och sen välja den bästa modellen med hjälp av korsvalidering. Kod till delar av denna uppgift finns här: trucks_cv.R.
 - (a) Skapa ett lämpligt tsibble-objekt för data, kalla objektet `trucks`. Skapa sedan data för korsvalidering enligt:

```
trucks_cv_tr <- trucks |>
stretch_tsibble(.init = 280, .step = 1)
```

Vad händer i koden ovan?
 - (b) Utgå från att vi vill göra prognoser för stegen $h = 1, 2, \dots, 11, 12$. Skatta modellerna med hjälp av korsvalidering. Notera: beräkningar kan ta lång tid. Vid behov kan ni testa att öka värdet `.init = 280` något i koden ovan.
 - i. SARIMA(1,1,1)(0,1,0) med en säsong på period 12
 - ii. SARIMA(3,1,0)(0,1,1) med en säsong på period 12
 - iii. SARIMA(3,1,0)(0,1,3) med en säsong på period 12
 - iv. En SARIMA som ni väljer själva.
 - (c) Ta fram en plott där ni på x-axeln har prognoslängd (kalla den `h`) och på y-axeln har RMSE. Varje modell ska bilda en sin egen linje med egen färg. Ta fram motsvarande plot, men med MAPE på y-axeln. Analysera och utvärdera. Vilken modell verkar bäst?
 - (d) Testa att anpassa en eller två exponential smoothing-modeller med hjälp av korsvalidering. Lägg in linjer för dessa modeller i plotten i (c). Jämför och analysera.
 - (e) Vilken modell anser ni vara bäst av de som ni har skattat? Ta den modellen¹ och skatta den på all tillgänglig data. Sen gör ni prognos för 12 steg framåt i framtiden. Skapa en plot med tidserien, prognosens och prediktionsintervall för prognosens.
5. FPP3 kap 13.8: Forecasting on training and test sets - Gå igenom och återskapa koden.

¹"Modellen" betyder här modellklassen och den övergripande strukturen, och inte specifika parametervärden. Dessa erhåller ni genom skattningen på data.

- (a) Gå igenom denna kod: länk Här visas det hur man kan göra för att skatta en modell på träningsdata och sen utvärdera dess prognoser på testdata² utan att skatta om modellen när vi flyttar startpunkten för prognoserna framåt.
- 6. FPP3 kap 10: Dynamic regression models, Gå igenom och återskapa koden i 10.1-10.6
- 7. FPP3 kap 12: Gå igenom och återskapa koden i 12.1 Complex seasonality
- 8. FPP3 kap 13: Gå igenom och återskapa koden i 13.1 Weekly, daily and sub-daily data
- 9. FPP3 kap 13: Gå igenom och återskapa koden i 13.9 Dealing with outliers and missing values

²Denna kod användes för att skapa figuren som visas i en slide på föreläsning 8.