# Labo 7: Stationaire tijdreeksen

## Doelstellingen

Een tijdreeks is een sequentie van waarden die door een proces in de tijd is gegeneerd (de onafhankelijk veranderlijke is namelijk de tijd). Voorbeelden zijn verkoopcijfers per kwartaal, buitentemperatuur per dag, wekelijks aantal bezoekers van een pretpark, geluidsignalen, hersengolven, het aantal actieve twittergebruikers maand na maand, enzoverder.

Het spreekt voor zich dat tijdreeksen bijzonder bruikbaar zijn in het analyseren van processen en het voorspellen van toekomstige uitkomsten. Op basis van patronen in historische gegevens kan men namelijk met een zekere betrouwbaarheid iets zeggen over de toekomst. Zo kan je op basis van de temperaturen van de afgelopen dagen min of meer inschatten wat de temperatuur van morgen of overmorgen zal zijn.

In dit labo zullen we ons beperken tot stationaire tijdreeksen. Typisch voor stationaire tijdreekseen is het feit dat de tijd geen invloed heeft op de statistische eigenschappen ervan. Zo zal het gemiddelde (variantie, correlatie, …) niet veranderen in functie van de tijd. Met andere woorden is er geen trend waarneembaar in de data waardoor de trendlijn een horizontaal verloop kent.

In de praktijk komen deze stationaire tijdreeksen in mindere mate voor. Veelal bevatten tijdreeksen namelijk wel een bepaalde trend. Denk bijvoorbeeld aan de wereldbevolking die jaar na jaar toeneemt, de aarde die stelselmatig opwarmt, prijzen die stelselmatig hoger worden, enzoverder.

Wel zullen we in dit labo een aantal basistechnieken zien om tijdreeksen te analyseren, signalen te filteren en voorspellingen te doen (time series forecasting). In het volgende labo werken we hierop verder door ook niet-stationaire tijdreeksen onder de loepe te nemen en dieper in te gaan op seizoenseffecten die zich kunnen voordoen.

## Inhoud

* Grafisch nagaan of een tijdreeks al dan niet stationair is
* Autocorrelatie
* Moving average
* Exponentiële demping
* Signalen als tijdreeksen
* Evalueren van voorspellingsmethodes

## Voorbeeld

Het tabblad “bezoekers” stelt het dagelijks aantal bezoekers (in honderden) voor een bepaalde attractie in een pretpark.

* We starten met het grafisch voorstellen van deze tijdreeks
* Via de trendlijn bepalen we of de tijdreeks al dan niet stationair is
* Via autocorrelatie kunnen we repetitieve patronen herkennen.
* Toepassen van onderstaande voorspellingstechnieken:
  + Voorspelde waarde is de huidige waarde
  + Voorspelde waarde is het 3-punts moving average
  + Voorspelde waarde is het 5-punts moving average
  + Voorspelde waarde via exponentiële demping waarbij we alfa kunnen variëren.
* Evaluatie van de voorspellingstechnieken via MAE (Mean Absolute Error) en de MAPE (Mean Absolute Percentage Error).

## Opdracht

#### Vraag 1 - Ziektegevallen

Een bedrijf houdt gedurende 90 dagen bij hoeveel van zijn werknemers zich ziek hebben gemeld.

De resultaten zijn te vinden onder het tabblad “zieken”.

1. Stel de data voor met een passende grafiek die duidelijk het verloop van het aantal zieken per dag weergeeft.
2. Ga na via de trendlijn of er al dan niet sprake is van een stationaire tijdreeks.
3. Kan je op basis van de grafiek terugkerende patronenen ontdekken?
4. Gebruik autocorrelatie om patronen te detecteren. Is er periodiciteit? Met andere woorden zijn er patronen die zich constant herhalen? Zo ja, hoe lang duurt zo een patroon?
5. Vergelijk onderstaande voorspellingsmethodes in termen van de MAE (Mean Absolute Error) en de MAPE (Mean Absolute Percentage Error).
   * Voorspelde waarde is de huidige waarde
   * Voorspelde waarde is het 3 punts moving average
   * Voorspelde waarde is het 5 punts moving average
   * Voorspelde waarde is het 10 punts moving average
   * Voorspelde waarde via exponentiële demping met alfa = 0,7
   * Zoek de optimale waarde van alfa gebruik makende van de Excel solver
6. Visualiseer de geteste methodes samen met het aantal zieken in 1 grafiek.
7. Welke methode geniet jouw voorkeur om mee te voorspellen? Argumenteer waarom.

#### Vraag 2 – Zonnepanelen

Het tabblad “zonnepanelen” bevat de data van een persoon die gedurende 2947 dagen bijgehouden heeft hoeveel elektriciteit zijn zonnepanelen hebben opgewekt en hoeveel ze hebben verbruikt.

1. Stel de data voor met een passende grafiek die duidelijk het verloop van de hoeveelheid elektriciteit opgewekt en verbruikt weergeeft.
2. Ga na via de trendlijn of er al dan niet sprake is van een stationaire tijdreeks.
3. Kan je op basis van de grafiek terugkerende patronen ontdekken?
4. Gebruik autocorrelatie om patronen te detecteren. Is er periodiciteit? Met andere woorden zijn er patronen die zich constant herhalen? Zo ja, hoelang duurt zo een patroon(ongeveer)?
5. Vergelijk onderstaande voorspellingsmethodes in termen van de MAE (Mean Absolute Error).

o Voorspelde waarde is de huidige waarde

o Voorspelde waarde is het 10 punts moving average

o Voorspelde waarde is het 20 punts moving average

o Voorspelde waarde is het 50 punts moving average

o Voorspelde waarde via exponentiële demping met alfa = 0,5

o Zoek de optimale waarde van alfa gebruik makende van de Excel solver

Visualiseer de geteste methodes samen met het aantal zieken in 1 grafiek.

1. Welke methode geniet jouw voorkeur om mee te voorspellen? Argumenteer waarom.

#### Vraag 3 – Gemiddelde temperatuur in Barcelona

Het tabblad temperaturen\_barcelona bevat de gemiddelde jaarlijkse temperatuur van Barcelona gemeten vanaf het jaar 1800 tot 1899

1. Ga na via de trendlijn of er al dan niet sprake is van een stationaire tijdreeks.
2. Gebruik autocorrelatie om patronen te detecteren. Is er periodiciteit? Met andere woorden zijn er patronen die zich constant herhalen? Zo ja, hoelang duurt zo een patroon(ongeveer)?
3. Zoek een optimale techniek voor het voorspellen van de volgende sample. Je kan kiezen uit alle technieken die tot nu toe aan bod zijn gekomen:

* Volgende sample is huidige sample
* Moving average met verschillende groottes
* Exponentiële demping. Zorg dat je de factor alfa kan aanpassen.

Evalueer telkens aan de hand van de MAE (Mean Absolute Error).

Hoeveel bedraagt de beste MAE die je hebt bekomen?

1. Visualiseer de gebruikte voorspellingstechnieken in een plot.
2. Argumenteer jouw keuze van voorspellingstechniek