

WETA907

Johdantoa aikasarja-analyysiin
*Introduction to Time Series
Analysis*

Timo J. Marjomäki
Jyväskylän yliopisto

- *Reading:* Chatfield, C. 1989: The analysis of time series: An introduction. Chapman & Hall.
- *Course* TILA220
- Aikasarjoja *Time series*
<http://www.robjhyndman.com/TSDL/>

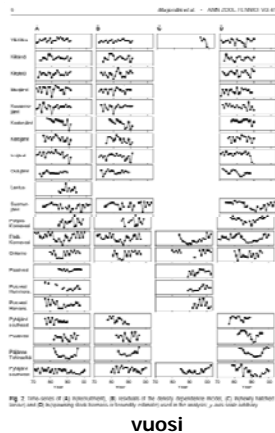
**Esimerkki:
populaatio-
dynamikka**

muikku
(*Coregonus albula*)

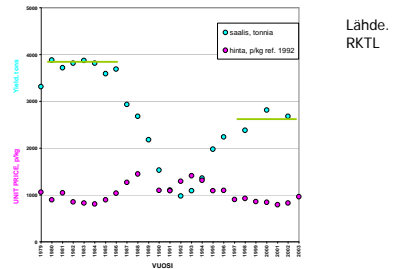
Runsauden vuotuiset indeksit

- A) $\ln(R)$
- B) $\ln(\text{residual of SB-R model})$
- C) $\ln(\text{larvae})$
- D) $\ln(\text{SB})$

where
 R = recruitment
 SB = spawning biomass



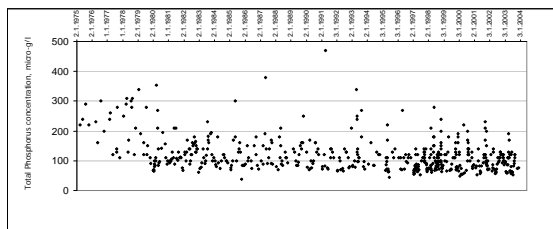
Esimerkki:
taloustiede *economics*
 Muikun vuosisaalis ja hinta (ammattikalastus)



Lähde:
 RKTL

see <http://www.economicswebinstitute.org/ecdata.htm> for economic series

Esimerkki:
Limnologia
 Kyrönjoen kokonaisfosforipitoisuus



Määritelmiä *Definitions*

- Aikasarja = ajallisesti järjestettyjen havaintojen joukko
- Miksi tarvitaan erityistä analyysiä
 - Sisältö: Havaintojen järjestyksellä voi olla merkitystä
 - Tilastotiede: havainnot eivät ole riippumattomia
- Kahdenlaisia aikasarjoja:
 - Jatkuvia (*continuous*)
 - Epä-jatkuvia (*discrete*)
- Kaksi vastakkaista syytä analysoida
 - Ilmiön ajallinen vaihtelu on kiinnostavaa
 - Esim. Lämpöäkö ilmasto todella
 - Halutaan poistaa ajallinen vaihtelu, jotta nähdään jotakin muuta
 - Esim. poistetaan vedenlaatu-aikasarjasta vuodenaikavaihtelu, jotta nähdään paremmin uuden puhdistamon vaikutus

Mitä aikasarjoilla voidaan tehdä

- Kuvata *Describe*: plotataan muuttujat ajan suhteen
 - Nähdään taso, trendi, syklit, poikkeavat havainnot ym.
- Selittää *Explain*: selitetään muuttujan muutosta ajalla
- Ennustaa *Forecast*: menneisyyden ja nykylukujen avulla arvataan tulevaisuutta
- Säädelä *Control*: yritetään vaikuttaa ilmiöön
- Tarkkailla *Monitor*: suunnitellaan ja toteutetaan

Muunnoksia, taas! *Transformations*

Tilastokäsittelyn vuoksi usein joudutaan tekemään havainnoille muunnoksia, esim. koska:

- Halutaan vakioida varianssi
 - Esim. jos $s.d.(x) \propto \text{average}(x) \rightarrow \ln(x)$ or \log
- Tehdään multiplikatiivisesta (kerrannaisesta) kausivaihtelusta additiivista (summautuvaa)
 - $x_t = m_t * s_t * \varepsilon_t \rightarrow$
 - $\log(x_t) = \log(m_t) + \log(s_t) + \log(\varepsilon_t)$
- Normalisoidaan satunnaisvaihtelu

Aikasarjan vaihtelun komponentit: *Components of variability*

1. Trendi (T) *Trend*
2. Kausivaihtelu (S) *Seasonal*
3. Sykli (C) *Cycle*
4. epäsäännöllinen (I), satunnainen, "virhevaihtelu" *irregular*

Kun 1-3 poistettu ja varianssi vakioitu = stationaarinen

1-3 removed and constant var -> stationarity

Aikasarjan variaation komponentit:

1. trendi *Trend*

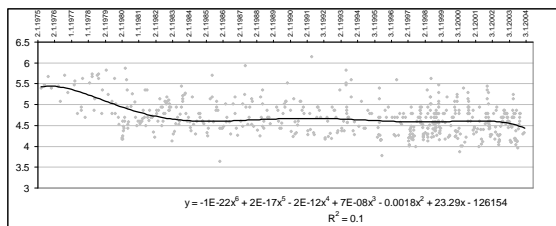
- *Trend is*
 - a **long term movement** in a time series, the underlying direction (an upward or downward tendency) and **rate of change** in a time series, when allowance has been made for the other components.
 - a general **systematic** linear or (most often) **nonlinear component that changes** over time and **does not repeat** or at least does not repeat within the time range captured by our data (e.g., a plateau followed by a period of exponential growth).

Aikasarjan variaation komponentit :

trendin estimointi regressiolla *estimating trend using regression*

- Suora *Line*
 $m_t = a + bt$
- Käyrä, esim. polynomi *Curve*
 $m_t = a + bt + ct^2 + \dots nb^p$

Esimerkki: ln(P), Kyrönjoki



EXCEL: kuvassa Right-click pisteiden päällä->valitse Add trend line->valitse Polynomial->valitse polynomin aste, tässä 6.

Aikasarjan variaation komponentit :
trendi

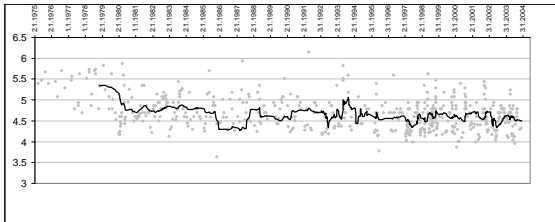
Filteröinti (*filtering of high frequency variability*)
pienentää vaihtelua, trendi näkyy selvemmin

- *MA=moving average*, liukuva keskiarvo
 - $y_t = (x_{t-q} + \dots + x_t + \dots + x_{t+q}) / (2q-1)$
 - Esim. Kausivaihtelun poisto, pituus s

SPSS: Transform->Create time series ->

- *Weighted smoothing*, painotettu
 - $y_t = (a_{t-q}x_{t-q} + \dots + a_t x_t + \dots + a_{t+q}x_{t+q}) / (2q-1)$
 - a=weight

Esimerkki:
ln(P), Kyrönjoki



EXCEL: in figure Right-click on any point->select Add trend line->select Moving average->give sequence length
Huom: EXCELissä perustuu aikaisempiin havaintoihin, tässä 25 kpl

Aikasarjan variaation komponentit :
trendin poisto differensoimalla
removing trend by differencing

- $Y_t = \nabla x_{t+1} = x_{t+1} - x_t$

SPSS: Transform->Create time series ->

- *E.g. linear trend with no seasonality*

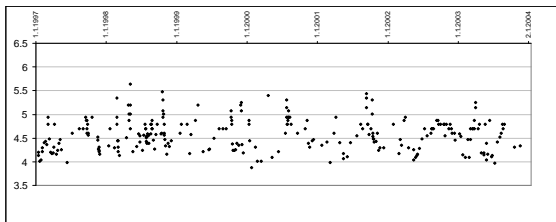
- $x_t = a + b \cdot t + \varepsilon_t$

- $\nabla x_{t+1} = (a + b \cdot (t+1) + \varepsilon_{t+1}) - (a + b \cdot t + \varepsilon_t) = b + \varepsilon_{t+1} - \varepsilon_t$

Aikasarjan variaation komponentit:
2. Kausivaihtelu *Seasonality*

- Vuoden-, kuukauden-, viikon- tai vuorokaudenajasta tms. johtuvaa
- Mikä tahansa (jakson sisäinen) säännöllinen vaihtelu
Any regular fluctuation
- *E.g. temperature outside: annual, daily*

Esimerkki:
ln(P), Kyrönjoki



Yleensä P-kuormitus vähäistä talvella ja suurinta syysateiden aikaan

Aikasarjan variaation komponentit :
kausivaihtelun estimointi
estimating seasonality

- S =kausivaihtelujakson pituus
- Estimointi:
 - Määrittele termit s_1, s_2, \dots, s_S
 - s_t =keskimääräinen poikkeama keskitasosta, esim. trendistä tai jos ei trendiä niin yleiskeskisarvosta
 s_t =average deviation from mean, trend etc.

Aikasarjan variaation komponentit :
kausivaihtelun poisto differensoinnilla
removing seasonality by differencing

- $y_t = \nabla_s x_t = x_t - x_{t-s}$
= $m_t - m_{t-s} + \varepsilon_t - \varepsilon_{t-s}$

Aikasarjan variaation komponentit :
3. Sykli *Cycle*

- Kuvaa säännöllistä (*regular*) vaihtelua
- Ei-kausittainen vaihtelu, jolla tietty jakso
- Ero kausivaihteluun: sykli yleensä pidempijaksoinen (ja jakson pituus voi vaihdella)

Aikasarjan variaation komponentit :
4. Epäsäännöllinen (*irregular*)

- Jää jäljelle kun edelliset poistetaan
- Joskus kiinnostavin
- Esim. Kyrönjoki
 - Kysymys: Aiheuttiko ruoppaus P-huhtoumaa
 - Menetelmä: Poistetaan trendi, kausivaihtelu ja ulkopuolisen muuttujan esim. sateen selittämä vaihtelu->jäljelle jäänyt vaihtelu voi johtua ruoppauksesta
 - Kontrollialueiden avulla osoitetaan, että epäsäännölliset muutokset leimallisia juuri ruopatulle alueelle (BACI)

**Aikasarjat ja korrelaatio:
autokorrelaatio
*autocorrelation***

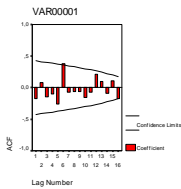
- r_k =autocorrelation with time lag k = correlation between observation at moment t and the observation at moment $t+k$
- Siis muuttujan virhetermit (residual error) tietyn ajan välein korreloivat
- Poistetaan trendi, kausivaihtelu jne.

**Aikasarjat ja korrelaatio
autokorrelaation merkitsevyys**

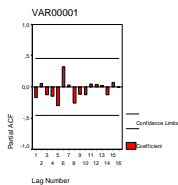
- random (non-autocorrelated) series of n observations $r_k \sim N(0, 1/n) \rightarrow$
 - 95 % of the estimates of correlation - $2/\sqrt{n} \dots 2/\sqrt{n}$
- Caution: if you calculate autocorrelations for several lags, then
 - Risk of finding at least 1 r_k significantly different from 0 increases risk $= 1 - 0.95^m$
 - m = number of estimated correlations
 - Bonferroni

**Aikasarjat ja korrelaatio
autokorrelaatio in SPSS**

Graphs->Time Series->Autocorrelations...



ACF: Correlates the values of a series with the values lagged by 1 or more cases. Autocorrelations are calculated for lags of 1, 2, ..., up to a specified number.



Partial ACF: Correlates the values of a series with the values lagged by 1 or more cases, after the effects of correlations at the intervening lags have been removed.

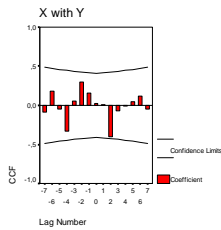
Aikasarjat ja korrelaatio kahden aikasarjamuuttujan ristikorrelaatio

cross-correlation (function) CCF

- Normaalisti muuttujien assosiaatio Pearson-korrelaatiolla
- Aikasarjoissa vaikutukset voivat olla viiveisiä
- $r_{xy}(k)$ = ristikorrelaatio viiveellä k = korrelaoidaan usean aikasarjan arvot viiveellä k , esim. Mja1:n tämän vuoden arvoja verrataan mja2:n ensi vuoden arvoon ($k=1$).
- Poistetaan trendi, kausivaihtelu ym.
- Merkitsevyyden kuten edellä

Aikasarjat ja korrelaatio *cross-correlation in SPSS*

Graphs->Time-Series->Cross-correlations...



Lyhyesti:

AR, MA, ARMA, ARIMA, SARIMA

- Erilaisia aikasarjalleja SPSS:ssä
 - autoregressiivinen malli (autoregressive models, AR)
 - liikkuvan keskiarvon malli (moving average, MA)
 - autoregressiivinen liikkuvan keskiarvon malli (ARMA)
 - autoregressiivinen integroiva liikkuvan (tai *liukuvan*) keskiarvon malli (ARIMA).

SPSS

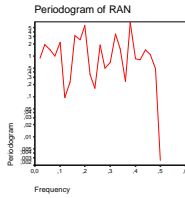
Analyze->Time Series->autoregression OR ARIMA

Lyhyesti:
spektrianalyysi
spectrum analysis

- Aikaulottuvuuden sijasta käsitellään taajuusulottuvuutta
- Aikasarja = eri taajuuksien "värähtelyiden" summa
 - $X_t = \mu + \sum_{j=1, \dots, k} R_j \cos(\omega_j t + \Theta_j) + Z_t$
 - ω (omega) = taajuus (frequency)
 - R = laajuus (amplitude)
 - Θ (theta) = vaihe (phase)
 - Z = stationaarinen satunnaissarja

Lyhyesti:
in SPSS

- *Graph -> Time Series -> Spectral*



*Periodogram:
Unsmoothed plot of spectral amplitude (plotted on a logarithmic scale) against either frequency or period. Low-frequency variation characterizes a smooth series. Variation spread evenly across all frequencies indicates "white noise."*

