

Norges befolkning

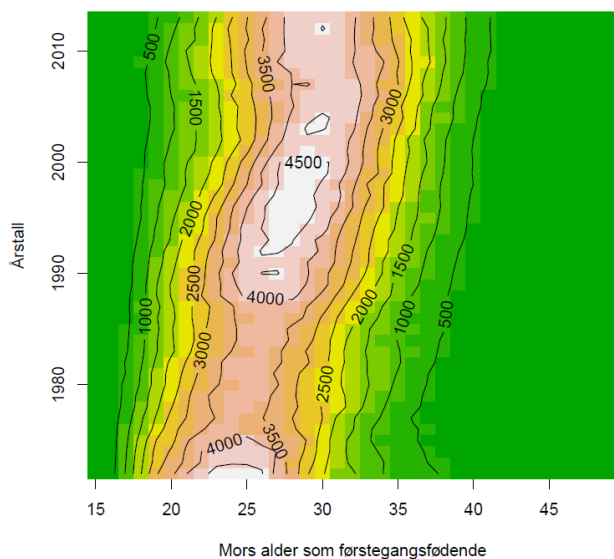
Nedenfor er det vist hvordan statistikkprogrammet og programmeringsmiljøet R (1) kan bli brukt til å analysere befolkningen i Norge.

Innhold

Mors alder ved fødsel av første barn.....	1
Menneskepopulasjonen i Norge.....	3
Barnefødsler i Norge.....	4
Befolkningseksplasjonen.....	11

Mors alder ved fødsel av første barn

Figur 1 og 2 viser at førstegangsfødende kvinner blir eldre.

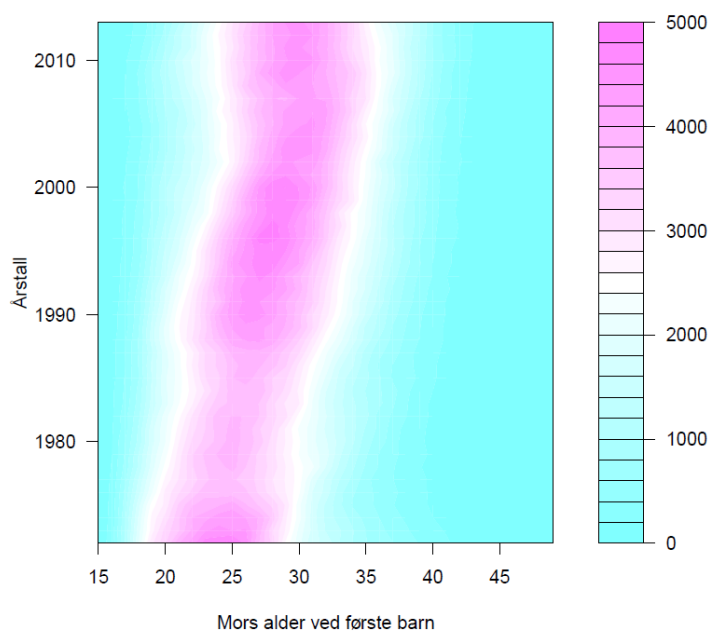


Figur 1 Mors alder (15-49 år) ved fødsel av første levendefødte barn i perioden 1972-2013. Konturene på figuren viser antall barn per år fordelt på årstall og mors alder. Gjennomsnittsalderen for førstegangsfødende har økt i perioden 1972-2005 for deretter å stabilisere seg.

Data SSB Tabell: 06990: Levendefødte, etter morens alder. Statistikkbanken

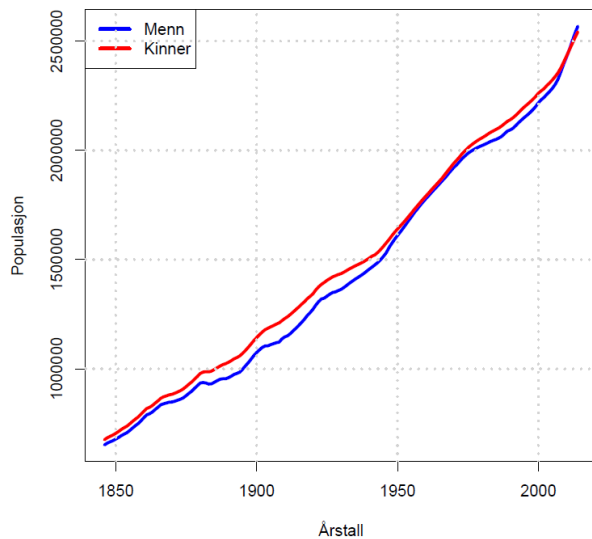
<http://www.ssb.no/>

Det er bra at kvinner skaffer seg utdanning og venter med å få barn. Dette er en trend i Europa, men dessverre skjer ikke dette på alle deler av kloden.



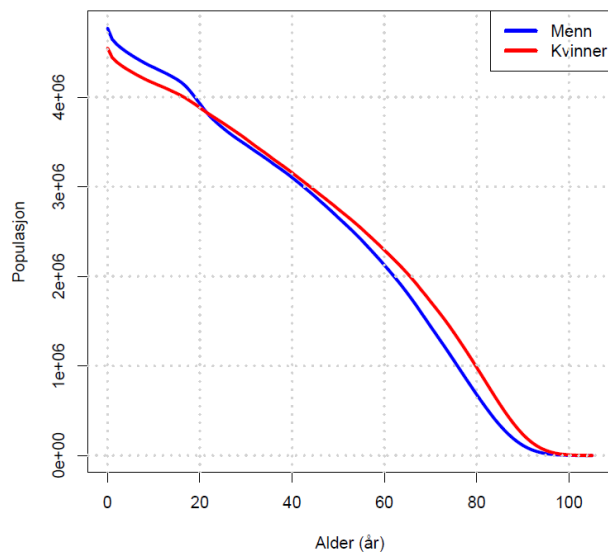
Figur 2 Konturplot som viser det samme som Figur 1. . Mors alder (15-49 år) ved fødsel av første levendefødte barn i perioden 1972-2013. Kotene på figuren viser antall barn per år fordelt på årstall og mors alder. Gjennomsnittsalderen for førstegangsfødende har økt i perioden 1972-2005 for deretter å stabilisere seg. Fargeskala fra turkis (færrest) via hvit til magenta (fleest). Data SSB Tabell: 06990: Levendefødte, etter morens alder.

Menneskepopulasjonen i Norge

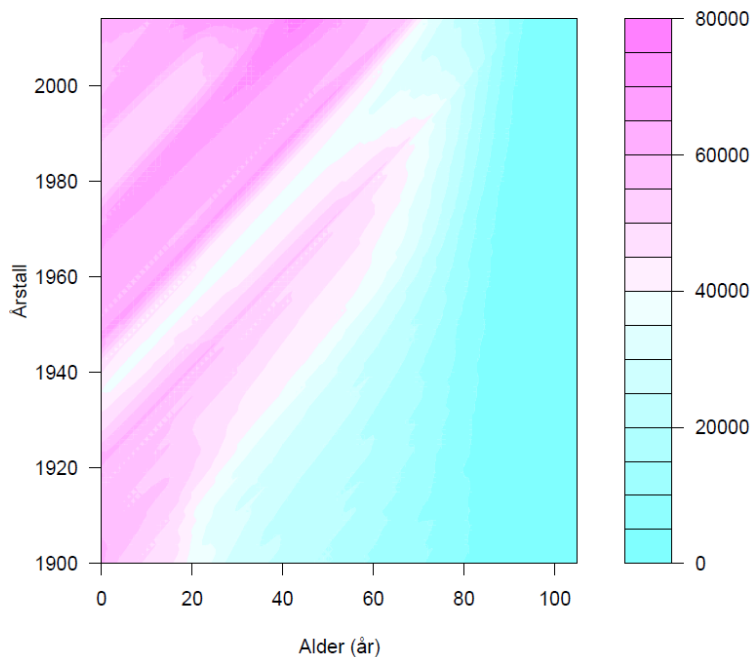


Figur 3 Befolkningen i Norge fra 1846-2014. I hele perioden har det vært flere kvinner enn menn, bortsett fra de siste årene hvor det skjer et skifte. Kurven har en skremmende stigning, og som begynner å stige brattere

Data: SSB/tabell: 10211: Folkemengde, etter kjønn og ettårig alder



Figur 4 Befolkningen i Norge fra 1846-2014 fordelt på aldersgrupper 0 til 105 år og mere. For alle aldersgrupper er det flere kvinner enn menn, bortsett fra de første ca. tyve leveårene hvor det er flest gutter. $1e+6 = 10^6 = 1$ million osv. Data: SSB/tabell: 10211: Folkemengde, etter kjønn og ettårig alder



Figur 5 Konturplot som viser folketallet (begge kjønn) i perioden 1900-2014 fordelt på alder og årstall. Legg merke til lave fødseltall fra i perioden 1930-1945, samt økning i etterkrigsgenerasjonen (sterk magenta) som nå etter hvert blir pensjonister og er en samfunnsutfordring i form av pensjonsutbetalinger og økende medisinske behov. Det var også en nedgang i antall fødseler fra 1980-1990 for deretter å øke igjen.

Data: SSB/tabell: 10211: Folkemengde, etter kjønn og ettårig alder

Barnefødsler i Norge

I tilknytning til artikkelen

Aarnes H & Andersen T : Barnefødsler 1967 – 2012 analysert i R

Tidsskr Nor Legeforen 2014; 134:2245 – 6

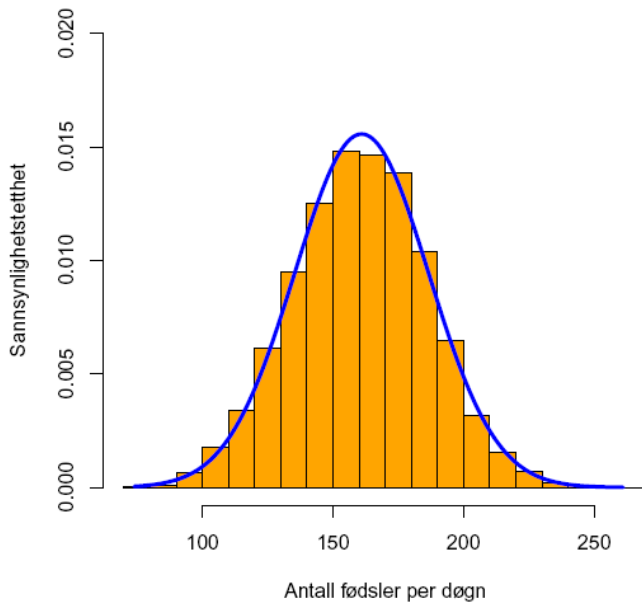
doi: 10.4045/tidsskr.14.1316

<http://tidsskriftet.no/article/3266554/>

var det flere figurer som det ikke ble plass til.

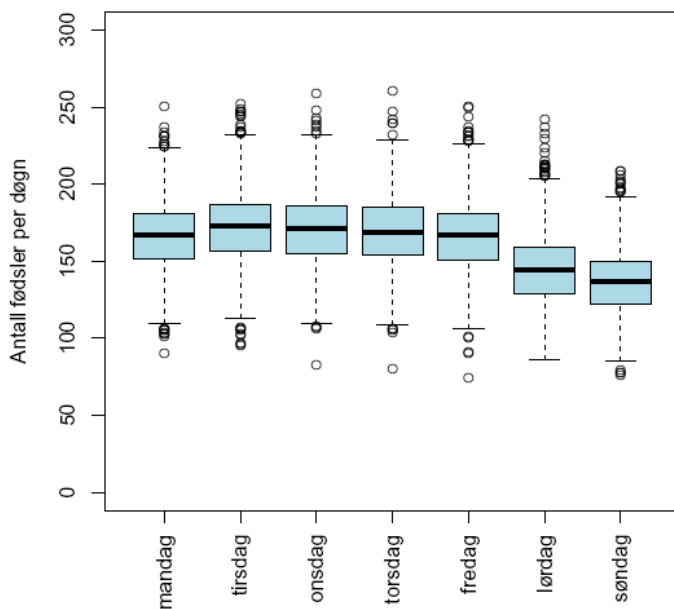
Tidsserieanalyse er studier av hvordan en variabel X utvikler seg over tid. **Trenden** viser utviklingen, mens **sesongvariasjonen** er typiske svingningsmønstre. I tillegg er det tilfeldig variasjon som ikke følger noe mønster. **Autokorrelasjon** er korrelasjon

mellom nærstående måleverdier i tidsserien. Tidsseriemodellen kan brukes til å lage prognoser om fremtidig utvikling.



Figur 6 Histogram av antall fødsler per døgn i Norge perioden 1967-2012. Den heltrukne linjen viser sannsynlighetstettheten for en normalfordeling. Data: Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt.

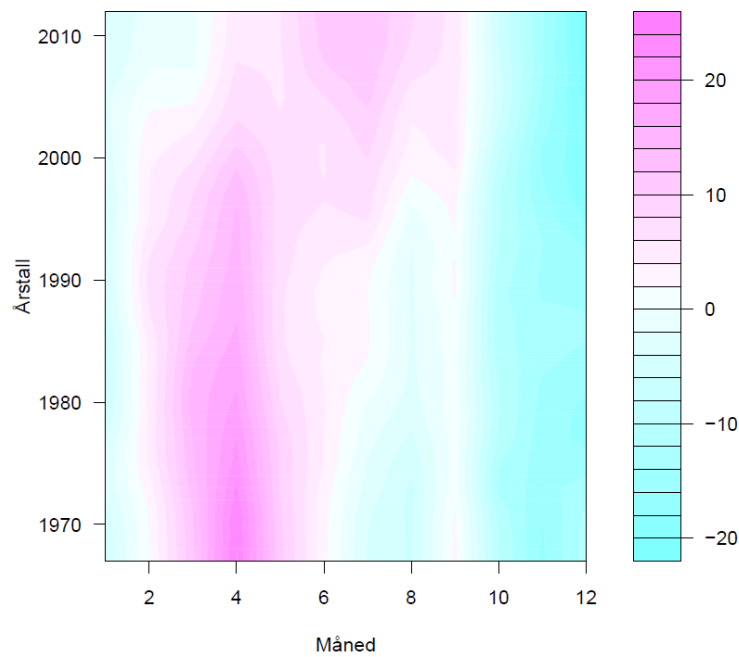
Figur xx viser at antall fødsler per døgn følger normalfordelingen.



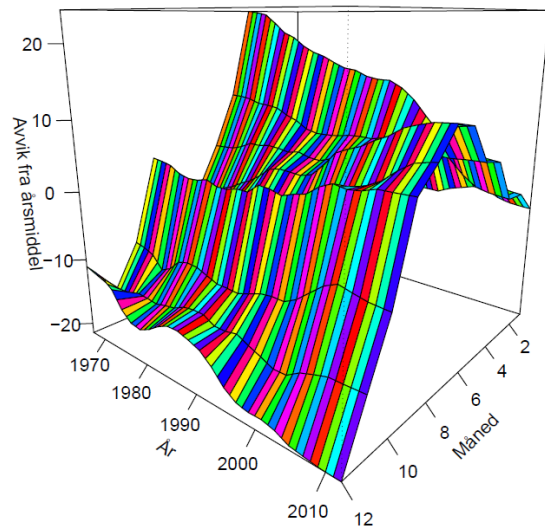
Figur 7 Boksplot viser antall barnefødsler per døgn i Norge perioden 1967-2012 fordelt på ukedag. Det er signifikant færrest fødsler i helgen (lørdag og søndag) sammenlignet med hverdager. Omfatter 46 år, 16802 døgn, 2401 mandager og søndager, og 2400 av de øvrige ukedagene, med i alt 2704646 fødsler, hvor antallet

varierer fra 74-261 fødsler per døgn. Data: Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt.

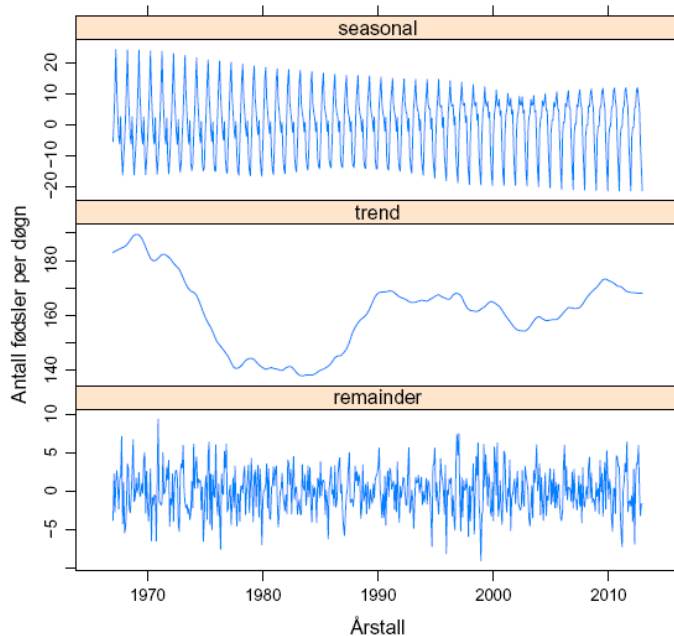
Figur 2 viser at det fødes omtrent like mange barn per døgn tirsdag til torsdag, noen færre mandag og fredag, men minst i helgen, lavest på søndager.



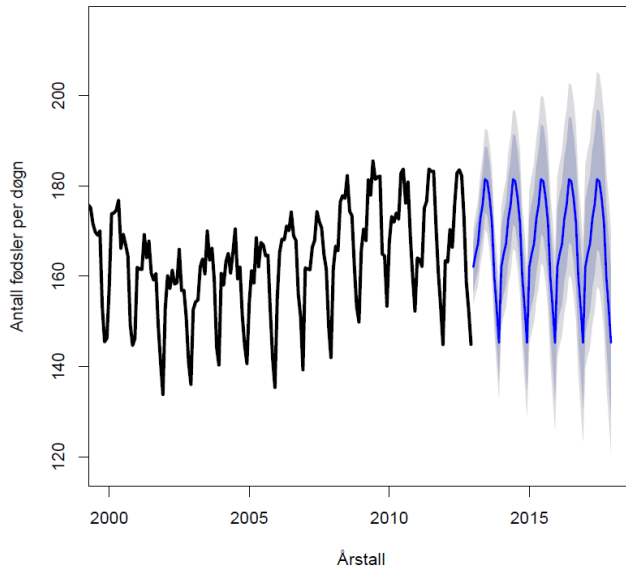
Figur 8 Sesongvariasjon i antall barnefødsler i Norge i perioden 1967-2012 vist som avvik fra årsmiddelerdien i form av en fargeskala fra turkis til hvit og magenta. Tallmatrisen for figuren inneholder 12 rader (måned) og 46 kolonner (årstall). Figuren viser trendforskyvningen i antall fødsler fra april til juli som har skjedd ved starten av 2000-tallet, og den fødselstoppen i september som har vart i 46 år. Fargen turkis (cyan) angir færre fødsler enn årsgjennomsnittet, et negativt avvik, og magenta flere fødsler enn årsgjennomsnittet, et positivt avvik. Hvit tilnærmet lik null avvik. Data: Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt.



Figur 9 Sesongvariasjon i antall barnefødsler i Norge i perioden 1967-2012 i et tredimensjonalt plot som avvik fra årsmiddeldverdien. Figuren viser det samme som fig 6 og 9, hvordan fødselstallet i oktober [10] til desember [12] har sunket, den stabile fødselstoppen i september [9] i hele perioden, samt den synkende trend for vårmånedene og en økning i sommermånedene fra starten av 2000-tallet. Data: Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt.



Figur 10 Tidsserieplot av antall fødsler i Norge 1967-2012 dekomponert i sesong, trend og en tilfeldig rest (remainder) når sesong og trend er trukket fra originalrekken. Et bevegelig gjennomsnitt brukes for å bestemme trend. Fra R-pakken lattice (2). Data: Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt.



Figur 11 ARIMA-modell. Ved automatisk ARIMA er den beste modellen ARIMA(3,1,4)(2,0,1) med periode 12. Figuren viser detaljutsnitt av tidsserien og trendfremskrivning ifølge ARIMA-modellen med 80% og 95% konfidensintervall. fra R-pakken forecast (3). Data: Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt.

Autokorrelasjon (seriell korrelasjon) er korrelasjon for en variabel med seg selv ved forskjellig tid.

I vanlig regresjon forutsetter man at restleddet (feilleddet) ϵ_t er normalfordelt og gjensidig uavhengig, hvit støy med lite informasjonsinnhold:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

I en tidsserie (tidsrekke) er ikke verdiene uavhengig av hverandre, de er **seriekorrelerete** eller **autokorrelerete**, og man kan ikke bruke vanlig regresjonsmodell. Vi kan beskrive korrelasjonen mellom den opprinnelige tidsrekken og en tidsforskjøvet rekke.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8

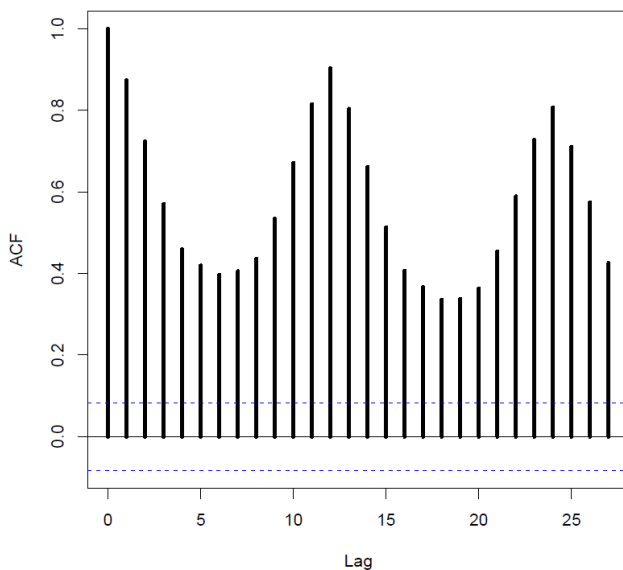
Den øverste rekken fra x_1 til x_t er den opprinnelig, og den under er forskjøvet med en tidsenhet, og man ser på korrelasjonskoeffisientene mellom disse to rekkene, kalt **autokorrelasjon lag 1**. Vi kan bruke et ledd i rekken x_t til å prediktere (forutsi) hva det neste leddet i rekken x_{t+1} vil bli. Hvis vi lager en tidsforskyvning på to ledd, **autokorrelasjon lag 2**, så sier dette noe om hvor god prediksjonen to tidsledd fram. Øverst en opprinnelige observasjonsrekken, nederst er rekken forflyttet to tidstrinn

frem, og på nytt ser vi på korrelasjonen mellom de leddene som står overfor hverandre

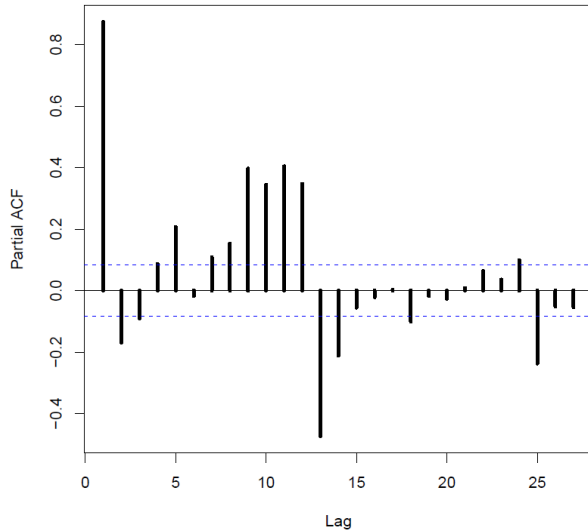
X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	
		X ₁	X ₂	X ₃	X ₄	X ₅	X ₆ X ₇

Slik kan man fortsette å forskyve og se på korrelasjonen mellom opprinnelig tidsrekke og forskjøvet tidsrekke, lag 3, lag4, ..., hvor måleenheten for lag nummer er måleenheten i tidsrekken.

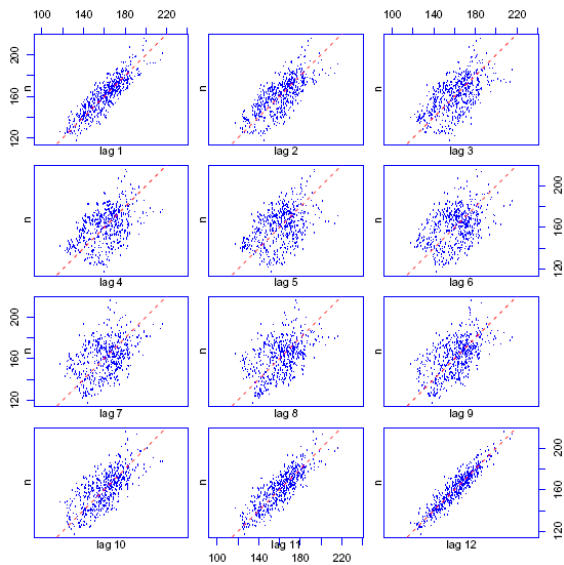
Hvis korrelasjonskoeffisienten mellom x_t og x_{t+1} er lik r og man antar at korrelasjonskoeffisienten mellom x_{t+2} og x_{t+2} også er lik r så vil korrelasjonen to trinn fram være r^2 . Imidlertid, hvis det er avvik mellom de to r i de to tidstrinnene så blir dette beskrevet av **partiell autkorrelasjon**, som angir avviket fra r^2 , og gir utfyllende informasjon om strukturen i rekken.



Figur 12 Korrellogram med autokorrelasjonsfunksjon (ACF) på y-aksen og lag nummer med enhet måned på x-aksen. Lag1 er en måned, lag2 er to måneder og lag12 er et år. Figuren viser en 12 måneders syklus. Korrelasjon 1 er perfekt korrelasjon. Den prikkete linjen angir ikke-signifikant korrelasjon. Laget i R-pakken forecast. Data: Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt.



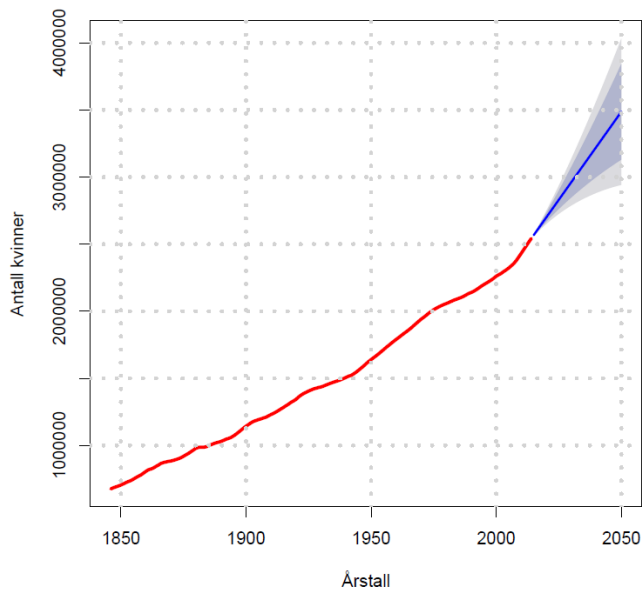
Figur 13 Korrelogram med partiell autokorrelasjonsfunksjon på y-aksen og lag nummer med enhet måned nummer på x-aksen. Figuren viser at etter den første positive autokorrelasjon ved lag 1 så blir lag 2 og lag 3 negativt korrelert, mens lag 5, og lag 7-12 er positivt korrelert. Den blå prikkete linjen viser partiell autokorrelasjon som ikke er signifikant forskjellig fra null. Fra R-pakken forecast. Data: Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt.



Figur 14 Lagplot av tidsserien lag 1 til lag 12. n er antall fødsler per døgn.

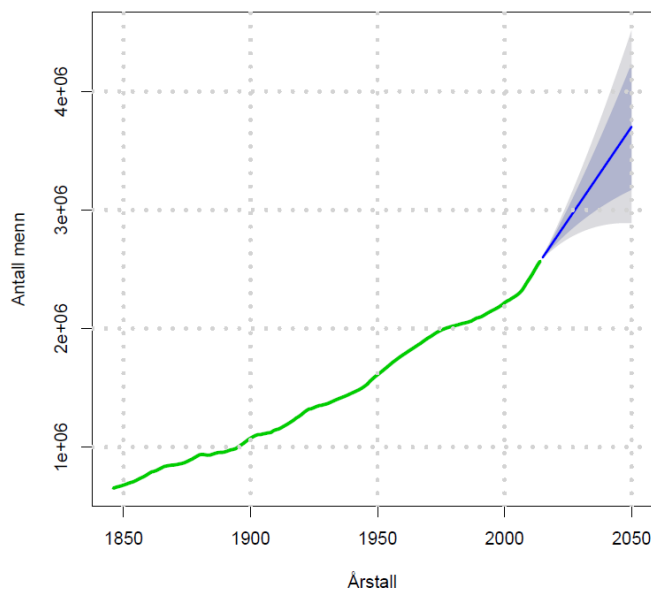
Befolkningseksplasjonen

Vi kan ved hjelp av R-pakken forecast (3) og bruk av en eksponentiell glattingsfunksjon (ETS) foreta en estimert framtidsbeskrivningen av populasjonen av kvinner og menn i Norge fram til 2050.



Figur 15 En framtidsbeskrivning av antall kvinner i Norge vil i 2050 være ca. 3.5 millioner angitt med 80% (lys grå) og 95% konfidensintervall (mørk grå).

Data: SSB/tabell: 10211: Folkemengde, etter kjønn og ettårig alder.



Figur 16 En framtidsbeskrivning av antall kvinner i Norge vil i 2050 være ca. 3.7 millioner angitt med 80% (lys grå) og 95% konfidensintervall (mørk grå).

Data: SSB/tabell: 10211: Folkemengde, etter kjønn og ettårig alder.

Til sammen antyder Figur 15 og 16 at Norges befolkningen i år 2050 vil være ca. 7.2 millioner.

Et problem er at politikere ikke forstår rekkevidden av eksponential vekst. Det velkjente eksemplet med 1 korn på første rute på sjakkbrettet, 2 på andre rute, 4 på tredje rute osv. Det stiger, men ikke urovekkende før man har passert halve sjakkbrettet, men så....

Gi verdens kvinner frihet, en god og lang utdanning samt inntektsgivende arbeid så vil automatisk fødselshyppigheten gå ned.

http://en.wikipedia.org/wiki/World_population

Den sjette masseutryddelsen

Det har gjennom de geologiske tidsperiodene vært fem store masseutryddelser av livet på Jorden (Ordovicium-Silur, sein Devon, Perm, Trias-Jura og Kritt-Tertiær).

http://en.wikipedia.org/wiki/Extinction_event

Mennesket sørger nå, i Holocen, for den sjette masseutryddelsen av Jordens arter, og sett i geologisk tidsskala med en skremmende hastighet.

Referanser

(1) R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

URL <http://www.R-project.org/>

(2) Sarkar, Deepayan (2008) Lattice: Multivariate Data Visualization with R. Springer, New York. ISBN 978-0-387-75968-5

(3) Hyndman RJ (2015). *forecast: Forecasting functions for time series and linear models*. R package version 5.8, <http://github.com/robjhyndman/forecast>.

HAA/2015