

Barnefødsler 1967–2012 analysert i R

R er et fritt tilgjengelig statistikk- og programmeringsspråk som blant annet kan benyttes til avansert simulering, modellering og grafisk presentasjon av data. Her viser vi hvordan R kan brukes til å analysere de over to millioner barnefødslerne i Norge i perioden 1967–2012. Vi påviser en faseforskyvning i antall fødsler fra vår til sommer, spesielt utpreget ved starten av 2000-årene.

R har åpen kildekode og kan lastes ned gratis (1). Det finnes en mengde tilleggspakker i R. Antallet er nå ca. 6 000, og stadig økende (2). Bioconductor, med ca. 930 programpakker basert på R, er et programmeringsmiljø i bioinformatikk og verktøy for analyse av alle typer genomiske data (3). R er et kommandolinjebasert skript-språk som kan benyttes innen alle former for statistikk og kvantitative beregninger, fra den enkleste form for kalkulator til avansert simulering og modellering, inkludert grafisk presentasjon av data.

Her viser vi et eksempel på R brukt i tids-serieanalyse av et datasett fra Medisinsk fødselsregister, Nasjonalt folkehelseinstitutt, som inneholder antall barnefødsler i Norge per døgn i perioden 1967–2012, i alt 2 704 646 fødsler. En tidsrekke eller tidsrekke av data er ordnet sekvensielt med faste tidsintervall der nærliggende datapunkter ofte er mer lik hverandre enn det man forventer ut fra ren tilfeldighet, det vil si at de er autokorrelerte. Månedsmiddelverdiene for daglig fødselstall har en karakteristisk sesongmessig variasjon og oscillasjon (fig 1). Gjennomsnittlig daglig fødselsrate i Norge falt i 1970-årene, men økte

i 1980-årene. Færrest blir født i perioden oktober til januar, flest i april (fig 2), og deretter synker fødselstallet utover i året, avbrutt av en topp i september.

En barnefødsel skjer i de fleste tilfeller tilknyttet et sykehus, der arbeidsrutiner er styrt av helgevakter og nasjonale høytidsdager. Sesongvariasjonen i fødselstall er overlappet

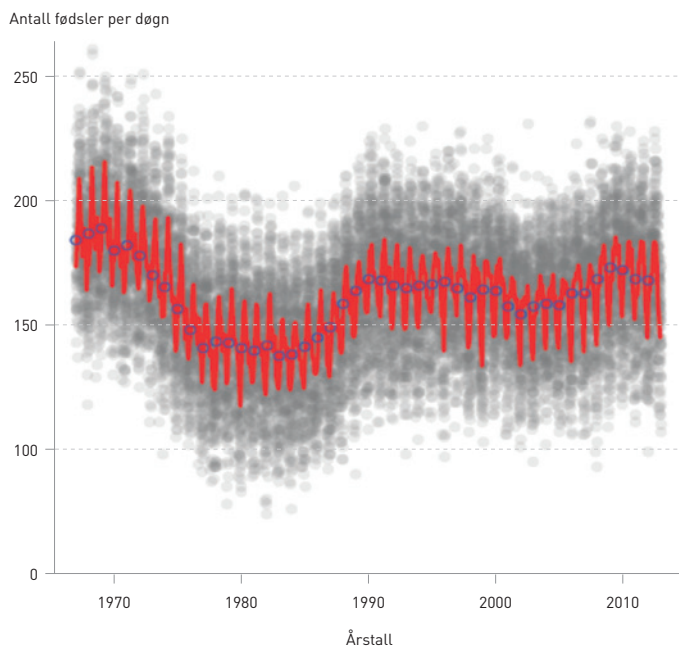
«Faseforskyvningen i antall fødsler fra vår til sommer var spesielt utpreget ved starten av 2000-årene»

med en tidstrend. Det er derfor nyttig å kunne fjerne tidstrenden for å kunne studere sesongvariasjonene for seg selv ved å dekomponere tidsrekken i trend, sesongvariasjon og uforklart restvariasjon (4, 5). Resultatet viser at selv om man fra 1967 til 2012 har hatt flest fødsler om våren, så skjer det en trendfor-

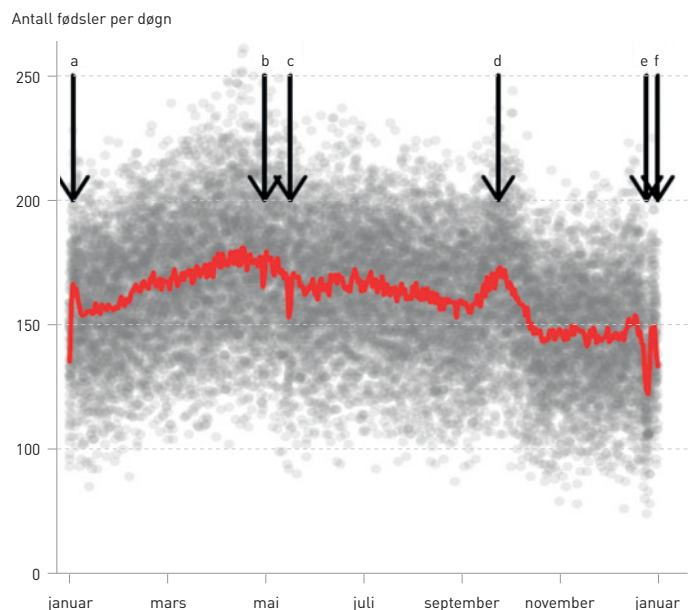
skyvning i sesongmønsteret, med en synkende trend for vårmånedene og en økning i sommermånedene sammenliknet med årsgjennomsnittet (fig 3). Figur 3 viser at trenden har vedvart hele perioden, mens figur 4 understreker at faseforskyvningen i antall fødsler fra vår til sommer var spesielt utpreget ved starten av 2000-årene.

Barnefødsler per døgn i Norge følger de samme hovedtrender og årstidsavhengige sykliske svingninger som man finner i andre deler av Europa (6). De fleste land har høyest fødselsrate om våren, og mange land har en septembertopp. Men hva er årsaken til sesongmønsterforskyvningen fra vår til sommer vi her observerer samt årstidssvingningene og toppen i september? Septembertoppen har sannsynligvis en enkel forklaring, ettersom den kommer ca. 38 uker etter jule- og nyttårsheglen, mens faseforskyvningen er vanskeligere å forklare.

I den tempererte sonen synkroniserer daglengden de biologiske døgn – og årstidsrytmene via lysreseptorer og hormonregulerte signalveier. Mennesket er seksuelt aktivt hele året, så hvis det ikke var andre ytre faktorer som påvirket, skulle man ikke forvente slike sesongmessige fluktuasjoner

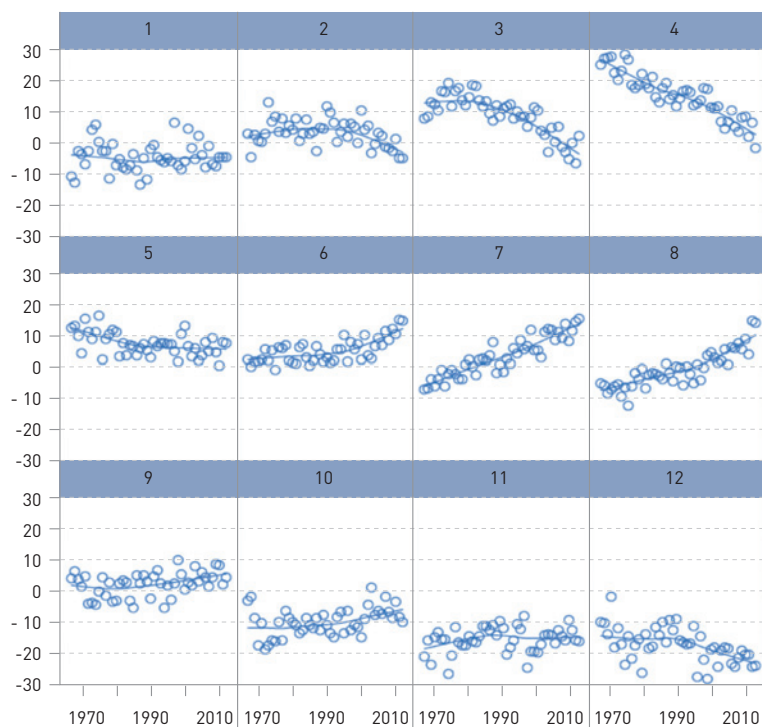


Figur 1 Antall barnefødsler per døgn i Norge i perioden 1967–2012. Den heltrukne linjen viser gjennomsnittsverdien per måned og de sykliske svingningene i fødselstall i løpet av året. De blå åpne sirkelene viser gjennomsnittsverdien per år. Amplituden er relativt konstant. Enkeltobservasjonene er gitt en gjennomskinnelig gråtone for å markere tydeligere overlapping mellom punktene

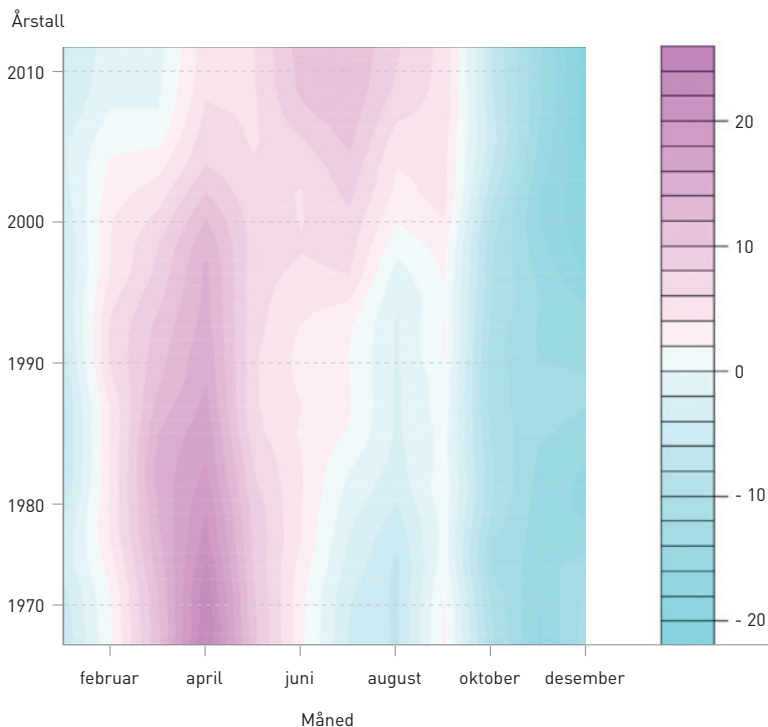


Figur 2 Antall barnefødsler i Norge i perioden 1967–2012 fordelt på månedene gjennom et år. Den heltrukne kurven viser gjennomsnittsverdiene. Pilene markerer ikke-bevegelige helligdager i Norge: a = 1. januar, b = 1. mai, c = 17. mai, d = septembertopp, e = 24. desember, f = 31. desember

Avvik fra årsmiddel (antall fødsler per døgn)



Figur 3 Sesongvariasjon i antall barnefødsler per døgn i Norge i perioden 1967–2012 som avvik fra årsmiddelverdien, gruppert månedvis fra januar [boks 1] til desember [boks 12]



Figur 4 Sesongvariasjon i antall barnefødsler i Norge i perioden 1967–2012 vist som avvik fra årsmiddelverdien i form av en fargeskala fra turkis via hvit til magenta. Fargen turkis (cyan) angir færre fødsler enn årsgjennomsnittet, et negativt avvik, og magenta flere fødsler enn årsgjennomsnittet, et positivt avvik, mens hvit angir tilnærmet lik null avvik

i fødselstall som man her observerer. Menneskets utvikling er styrt av biologisk og sosiokulturell evolusjon, og man kan forvente å finne spor av begge i reproduksjonsmønsteret.

I preindustrielle samfunn hadde biologiske og abiotiske faktorer stor betydning for reproduksjonen. Blant de sosiokulturelle faktorene som kan påvirke tidspunkt for en barnefødsel er utdanning, foreldrepenger og lønnsinntekt (unge uten lønnsarbeid får en lav engangsstøtte), global migrasjon, urbanisering, religiøse helligdager, ferievaner med økt soling, samleiefrekvens, bryllups-sesong, økt alder hos førstegangsfødende, fødselspermisjon og kontantstøtteordning for småbarnsforeldre innført i 1998. Det er naturlig å tenke seg at de sykliske svingningene i fertilitet, unnfangelse og fødselstidspunkt hos mennesker er reminisenser fra vår biologiske evolusjonære forhistorie og livshistoriestrategi. Faseforskyvningen i sesongmønsteret for fødselsraten i Norge kan derfor antakeligvis ikke bare tilskrives kontantstøtteordningen.

Slik dette eksempelet viser, kan R brukes til å analysere tidsrekker (f.eks. funksjonell MR, EEG, trykksignaler og kroppsbevegelse). R kan utveksle data med de fleste andre statistikkpakker og kan anvendes innen alt fra epidemiologi, overlevelsesanalyse og SIR-modeller (Susceptible-Infected-Recovered-Model) til metagenomikk.

Halvor Aarnes

halvor.aarnes@ibv.uio.no

Tom Andersen

Tom Andersen (f. 1955) er professor i limnologi ved Institutt for biovitenskap, Universitetet i Oslo.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Halvor Aarnes (f. 1948) er professor i plantefysiologi ved Institutt for biovitenskap, Universitetet i Oslo.

Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Litteratur

1. The R Project for Statistical Computing. www.R-project.org/ [31.10.2014].
2. The Comprehensive R Archive Network. <http://cran.r-project.org/> [31.10.2014].
3. Bioconductor. Open source software for bioinformatics. www.bioconductor.org/ [31.10.2014].
4. McLeod AI, Yu H, Mahdi E. Time series analysis with R. Handbook of Statistics 30. Cambridge: Elsevier, 2011: 1–75.
5. Metcalfe AV, Cowpertwait PSP. Introductory time series with R. Heidelberg: Springer, 2009.
6. Bronson FH. Seasonal variation in human reproduction: environmental factors. Q Rev Biol 1995; 70: 141–64.

Mottatt 22.10. 2014, første revisjon innsendt 11.11. 2014, godkjent 14.11. 2014. Redaktør: Are Brean.

Publisert først på nett.