

Vitenskap, tallmagi og kritisk sans

Hvordan kan tallmagi bli fakta? Verken redaksjoner eller lesere bør legge til side egen fagkunnskap og kritisk sans, selv når det refereres til avanserte statistiske fordelinger med tilhørende egennavn, og materialet tilsynelatende er godkjent av statistikere.

Liv Ariane Augestad
l.a.augestad@medisin.uio.no
Mathias Barra
Samantha Salvesen Adams

Følgende er sakset fra spalten *Fra andre tidsskrifter* i Tidsskriftet nr. 19/2014 (1): «Benfords lov, også kalt førstesifferloven, sier at førstesifferet 1–9 for tall i naturlige vekstprosesser er høyest for 1 og lavest for 9. Man har lenge antatt at fødsler er likt fordelt på månedens datoer, men i en nylig publisert artikkel i tidsskriftet *Tilfeldig gang*, utgitt av Norsk statistisk forening, viser [T. Dønvold] at Benfords lov også gjelder for fødsler [(2, 3) ...] Som en regel fødes det altså flere jo lavere tverrsummen av dagnummeret er, dvs. flest fødsler når tverrsummen er 1 og færrest når den er 9. Det betyr igjen flest fødsler den 1., 10., 19. og 28. hver måned og færrest den 9., 18. og 27. [...] Studien kan bidra til bedre bemanningsplanlegging på fødselsklinikker.»

Forfatterne av denne kronikken var kort tid etter oppslaget samlet i et middagsselskap, og samtalen kom raskt inn på dette funnet. Samantha Salvesen Adams og Liv Ariane Augestad var forvirret. De var tidligere ukjent med Benfords lov, men syntes likevel det var underlig at fødsler skulle være hyppigere på datotall med lav tverrsum. Originalartikkelen i *Tilfeldig gang* (2) ga heller ingen god forklaring. Hvilken mekanisme kunne ligge til grunn for at antall fødsler per dag har en syklisk fordeling gjennom måneden, tilsynelatende styrt av 10-tallssystemet? Effekten skulle være så betydelig at det kunne ha konsekvenser for bemanningsplanleggingen på fødselsklinikker. Til andre middagsgjesters mulige forkleinelse brukte vi en god del av selskapet til å oppklare hva Benfords lov er, og til å legge hodene i bløt for å finne en mulig sammenheng med fødsler.

Hva vet vi fra tidligere?

Fødsler utviser sesongvariasjon som har med sesongvariasjon i befruktning å gjøre, relatert til ferie og fridager og sannsynligvis også til tidsfrister for barnehageopptak (3, 4). Det er også rapportert om systematisk

ukevariasjon med færre fødsler i helgene, spesielt søndager (5, 6). Dette er relatert til samfunnsorganiseringen – færre planlagte keisersnitt, mindre intensiv fødselsomsorg, som indusering av fødsler og muligens også relatert til gravide kvinners liv i helgene. Kunne det være noe tilsvarende rundt datoer med lav tverrsum, altså et underliggende samfunnsfenomen som ville kunne påvirke de gravidenes livsførsel eller fødselsomsorgens intensitet etter datotallenes tverrsum?

Benford

Mathias Barra hadde ikke lest innlegget, men kunne bidra med matematisk kompetanse og en viss grad av kjennskap til Benford. Kunne en forklaring på Benfords lov gi teoretisk støtte for en Benford-fordeling av fødsler?

Kort fortalt sier vi at Benfords lov holder for en tallfølge dersom fordelingen av førstesifre i tallfølgen er såkalt Benford-fordelt. Benfords lov holder typisk for tallfølger av måltall, som er tall som kvantifiserer en eller annen størrelse. Et eksempel på en måltallfølge som – antakelig tilnærmet – oppfyller Benfords lov er lengdene til alle verdens elver. Målt i kilometer vil det være flest korte elver og bekker av lengde 1,00–1,99 kilometer. Disse har alle et førstesiffer lik 1. Det er litt færre av lengde 2,00–2,99 kilometer som har førstesiffer 2, og enda færre av lengde 9,00–9,99 kilometer med førstesiffer i måltallet lik 9. Generelt, på hvert intervall fra og med 10^n frem til 10^{n+1} km, er det en tendens til at elvene med lengde i dette intervallet er flest i den første delen, der hvor førstesifferet er 1.

Dersom man bruker 10-tallssystemet, vil man se en spesifikk Benford-fordeling av førstesifrene – ca. 30 % 1-tall, fallende til ca. 5 % for 9-tall. Det finnes én Benford-fordeling for hvert grunntall, $b \geq 2$, og for at man skal kunne si at Benfords lov gjelder, skal førstesifrene være Benford-fordelte uavhengig av i hvilket tallsystem måltallet er representert, og uavhengig av hvilken måleenhet som er benyttet. Om vi i stedet måler elvene i fot eller tommer, velger oss binære tall eller desimaltall – dette fenomenet vil gå igjen grunnet det underliggende fenomenet: at tallene er måltall på noe

fysisk der det er flest små og færre og færre større. Den interesserte leser kan finne mer informasjon på Wikipedia (7).

Er det en god hypotese?

Etter intens idémyldring kunne vi ikke komme på noe i samfunnet som organiseres slik at det skulle gi høyere fødselstall på datoer med lav tverrsum. Videre har Benfords lov å gjøre med fordelingen av førstesifre i måltall. Datotall er ikke måltall. De er ordenstall og sier ingenting om størrelse på den aktuelle dato.

Om det empiriske datasettet skulle understøtte at det faktisk forekommer flere fødsler på datoer med lav tverrsum, ville fortsatt ikke Benfords lov gjelde, med mindre vi også fant det samme ved å dele tiden inn i uker eller timer, eller bestemte oss for å bruke et 5-tallssystem for månedens datoer. Vi opererer altså uansett ikke innenfor et domene der Benfords lov gjør seg gjeldende. Med denne grunnlagskunnskapen på plass var det naturlig å undersøke saken videre. Første skritt var å gå originalartikkelen nærmere etter i sømmene.

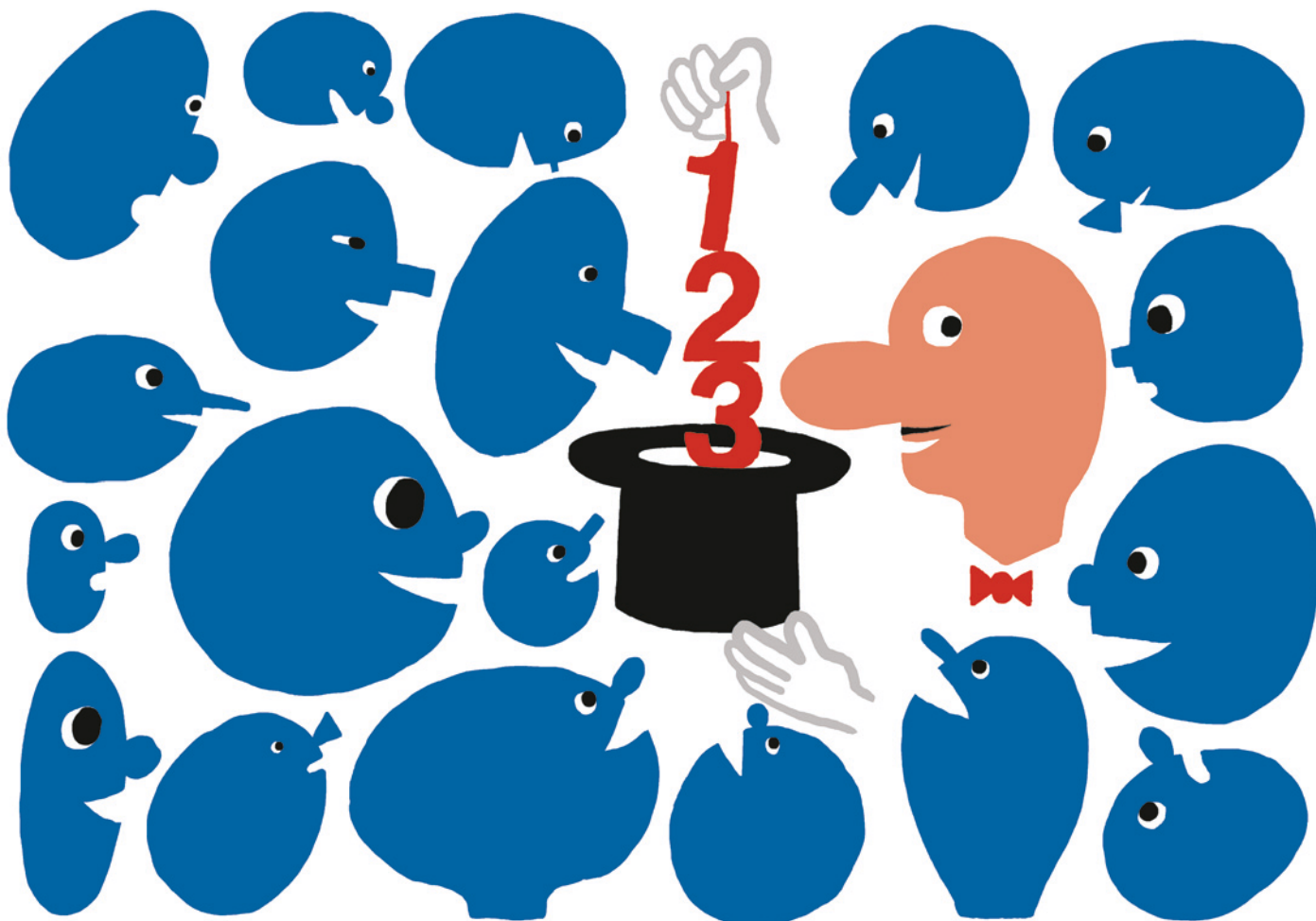
Kildekritikk

Da vi ringte redaksjonen til studenttidsskriftet *Tilfeldig gang*, kunne fungerende redaktør Turid Follestad umiddelbart bekrefte at *Tilfeldig gang* ikke er et fagfellevurdert tidsskrift, og at ingen i redaksjonen eller Norsk Statistisk Forening for øvrig ville gå god for innholdet i artikkelen (2). Fra studenttidsskriftets nettsider var det lenket til en lengre versjon av originalartikkelen som også inkluderte data (8). Lenken til fullversjonen fra hjemmesiden forsvant noen få dager etter at vi snakket med redaksjonen i *Tilfeldig gang*.

Støtter dataene konklusjonen?

Hypotesen fra artiklene (1, 2 og 8) kan dog tilbakevises uten tilgang til nyskapende kildekritisk teknikk som telefoni, ettersom det ikke presenteres noe gyldig matematisk argument. Datamaterialet fra fullversjonen (8) er offentlige data om amerikanske forsikringskunders registrerte fødselsdatoer, og analysene i originalartikkelen gir faktisk ingen støtte til konklusjonen om Benford-fordeling av fødsler.

Tvert imot er dataene presentert i artik-



Illustrasjon: Espen Friberg

lene, *ikke* Benford-fordelte. Enkle utregninger viser at hvert førstesiffer 1–9 opptrer med en frekvens på mellom 11,3 % og 10,9 %. Ikke akkurat noen overbevisende fallende tendens – med mindre man, som i de aktuelle artiklene, får dette forelagt i grafisk form med avkortet y-akse. De dataene som ikke passer med konklusjonen, gis en ad hoc-forklaring. Den endelige konklusjonen om at fødsler er Benford-fordelte på datoer står derfor i direkte motstrid til de dataene som faktisk presenteres.

Noe å lære av?

At denne «artikkelen» fremheves i spalten *Fra andre tidsskrifter* og dermed fremstår som fagfellevurdert, vitner om at redaksjonen har hatt en svikt i rutinene. Slikt kan skje den beste. Det er likevel ekstra graverende at det skjer i nettopp en slik sak som dette. Det er uforståelig for oss hvordan noen kjent teori skulle lede frem til hypotesen om Benford-fordeling av fødsler, og funn av en slik sammenheng ville revolusjonert vårt syn på hva et datotall representerer.

Vitenskapen har ikke svar på alt, og

iblant gir empirien støtte for sammenhenger som kan være vanskelig å forklare initialt. Det finnes flere historiske eksempler på dette. For oss fremstår likevel en hypotese om at fødsler skulle hope seg opp på datoer med lav tverrsum å være mer basert på en form for tallmagi enn kunnskap om fødsler eller Benfords lov.

Fagfellenes vurdering

Den rapporterte fødselsfordelingen er med andre ord såpass overraskende at redaksjonen burde mistenke ugler i mosen. Enkel kildekritikk ville ha avslørt at denne artikkelen ikke var en som Tidsskriftet burde bringe videre. I en vitenskapelig publiseringsverden der det finnes stadig flere dubiose «fagfellevurderte» tidsskrifter, stilles det økende krav til lesernes kritiske sans og evne til selv å vurdere innholdet i artikler. En artikkel publisert i et anerkjent tidsskrift er heller ingen garanti for at den faktisk er fagfellevurdert: Nylig har en rekke godt etablerte forlag trukket tilbake over 100 artikler på grunn av fagfellesvindel (9). Fagfellevurdering er i stor grad ubetalt

arbeid, og kvaliteten er sterkt varierende. Selv når det ikke er direkte svindel involvert, kan vurderingen være under enhver kritikk og utført av inkompetente likemannsdommere.

Fagfellene er heller ikke immune mot tallenes imponerende effekt: Å inkludere en matematisk formel høyner sjansene for at forskningen blir vurdert til å være av høy kvalitet, til og med hvis formelen er ureltert til den forskningen (10). Det finnes også fagfellevurderte tidsskrifter der det er grunn til å sette spørsmålsteget ved hele fagparadigmet. For eksempel huser store forlag som Elsevier også homøopatitidsskrifter, med fagfellevurdering og «impact factor», indeksert i Medline og med uttelling som nivå 1-tidsskrift her til lands.

Hvor viktig er hypotesen?

Hypoteser som testes empirisk innenfor naturvitenskap generelt, og i medisin spesielt, er ofte generert som en komponent i en teori som søker å forklare observasjoner. For eksempel kan kliniske spesialister danne en hypotese bygd på erfaring og kunnskap

om patofysiologi, som kan være en god kandidathypotese gitt hittidig kunnskapsbase. Historisk har det vært store praktiske begrensninger ved hvilke og hvor mange hypoteser det var mulig å undersøke eksperimentelt, og det er grunn til å anta at seleksjon av testede hypoteser har hatt en nær relasjon til hypoteser man har ment var mest sannsynlig riktige. Med økende datatilfang, regnekraft og publiseringspress kan vi anta at mer eller mindre tilfeldig testing av sammenhenger øker i jakten på signifikante p-verdier. Det stiller større krav til leserens kompetanse til å vurdere hvor godt forskerne begrunner testingen av en gitt hypotese.

Et tankeeksperiment

Anta at noen forteller deg at man i et stort datasett har funnet at 30-dagers overlevelse etter gastrokirurgi øker (oddsratio eller et annet statistisk mål) om den tredje bokstaven i fornavnet til kirurgen er «d», eksempelvis Lydia eller Anders. Videre er p-verdien til effekten oppgitt å være $p = 0,034$. Under nullhypotesen om at kirurgens tredje fornavnsbokstav ikke gir økt sannsynlighet for bedre overlevelse «er det (kun) 3,4 % sjansje for å observere en like skjev eller skjevtere fordeling av 30-dagers overlevelse». Betyr dette at det er 96,6 % sjansje for at den angivelige «d»-effekten er sann? Nei! Funnet sier ingenting direkte om sannsynligheten for at en hypotese er sann eller ikke sann. Det må tolkes i sammenheng med den tiltroen man har til hypotesene som testes. Her eksemplifisert ved nullhypotesen om at kirurgens tredje fornavnsbokstav ikke påvirker 30-dagers overlevelse – noe man bør ha en rimelig grei tro på. Denne tiltroen bør ikke svekkes betraktelig, selv med en såkalt signifikant p-verdi.

Da italienske forskere i 2011 «fant» par-tikler som beveget seg raskere enn lysets hastighet, ba de andre om hjelp til å finne feilen. Under nullhypotesen «relativitetsteorien er sann» hadde eksperimentene en p-verdi på 0,000002. Et stort internasjonalt samarbeid ble satt i gang, og flere måneder senere fant man til slutt en systematisk målefeil, og dermed står Einsteins relativitetsteori fortsatt fast (11).

Hva hvis analyser på et empirisk datasett faktisk viste en opphoping av fødsler på

dager med lav tverrsum, og at nullhypotesen om en jevn fordeling på datotall hadde en p-verdi lavere enn 0,05? Har vi en godt begrunnet (null)hypotese om at datotall ikke henger sammen med fødselsrater, burde vi ikke forkaste nullhypotesen. Det er mer sannsynlig at det empiriske datamaterialet er en tilfeldig observasjon av en underliggende jevn fordeling enn at det faktisk skulle være en sammenheng mellom datotall og antall fødsler. Jo mer teori som tilsier at nullhypotesen er riktig, desto sterkere statistiske indikasjoner trenger vi for å falsifisere den.

Når skepsisalarmen blinker

For det vitebegjærlige mennesket er det lite som er artigere enn ny og revolusjonerende kunnskap. Forskning krever nysgjerrige forskere og i mange tilfeller en klype kreativitet. God forskning krever imidlertid også kjennskap til etablert kunnskap og det man kan kalle sunn fornuft. Som en generell regel vil vi hevde at paradigmeskiftende konklusjoner i vitenskapelig arbeid bør ha meget overbevisende evidens. Også tidsskriftredaksjoner og lesere har mulighet til å konsultere sin kritiske sans og fagkunnskap når de vurderer nye resultater.

Hvis skepsisalarmen begynner å blinke, anbefaler vi særlig tidsskriftredaksjoner å foreta videre undersøkelser. «Keep an open mind, but not so open that your brain falls out» (12).

Liv Ariane Augestad (f. 1980)

er ph.d., lege og postdoktor ved Avdeling for helseledelse og helseøkonomi, Universitetet i Oslo og Helse Øst kompetansesenter for helsetjenesteforskning ved Akershus universitetssykehus. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Mathias Barra (f. 1977)

er ph.d. i matematisk logikk og forsker. Han arbeider med matematisk modellering av pasientforløp, helseøkonomi, statistikk og livskvalitetsforskning ved Helse Øst kompetansesenter for helsetjenesteforskning ved Akershus universitetssykehus. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Samantha Salvesen Adams (f. 1981)

er ph.d og fastlegevikar ved Frysja legekantor. Forfatter har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Litteratur

1. Dønvold T. Sesongjusterte fødsler er ikke jevnt fordelt på datoer. Tidsskr Nor Legeforen 2014; 134: 1834.
2. Dønvold T. Benfords lov i fødselsdata. 2014. Tilfeldig gang. <https://sites.google.com/site/statistiskforening/tilfeldig-gang> (31.8. 2014).
3. MacFarlane A. Variations in number of births and perinatal mortality by day of week in England and Wales. BMJ 1978; 2: 1670–3.
4. Brunborg H. Fødsler i jule- og nyttårshelgen: Få fødsler, mange unnfanger. SSB 2001. <http://ssb.no/befolkning/artikler-og-publikasjoner/fa-fodsler-mange-unnfanger> (20.11. 2014).
5. Gam CM, Tanniou J, Keiding N et al. A model for the distribution of daily number of births in obstetric clinics based on a descriptive retrospective study. BMJ Open 2013; 3: e002920.
6. Barra M, Lindstrøm J, Adams S et al. Sesongjusterte fødselsfrekvenser er poissonfordelte. Tidsskr Nor Legeforen 2015; 135
7. Benford's law. http://en.wikipedia.org/wiki/Benford's_law (1.6. 2015).
8. Dønvold T. Sesongjusterte fødsler er ikke uniformt fordelt. Tilfeldig gang 2015.
9. Ferguson C, Marcus A, Oransky I. Publishing: The peer-review scam. Nature 2014; 515: 480–2.
10. Eriksson K. The nonsense math effect. Judgm Decis Mak 2012; 7: 746–9.
11. Pressemeldinger fra CERN. OPERA experiment reports anomaly in flight time of neutrinos from CERN to Gran Sasso. <http://press.web.cern.ch/press-releases/2011/09/opera-experiment-reports-anomaly-flight-time-neutrinos-cern-gran-sasso> (30.5. 2015).
12. Quote investigator. <http://quoteinvestigator.com/2014/04/13/open-mind/> (30.5. 2015).

Mottatt 1.7. 2015 og godkjent 7.10. 2015. Redaktør: Hanne Støre Valeur.



Engelsk oversettelse på www.tidsskriftet.no