



Poissonfordelingen for antall hendelser

MEDISIN OG TALL

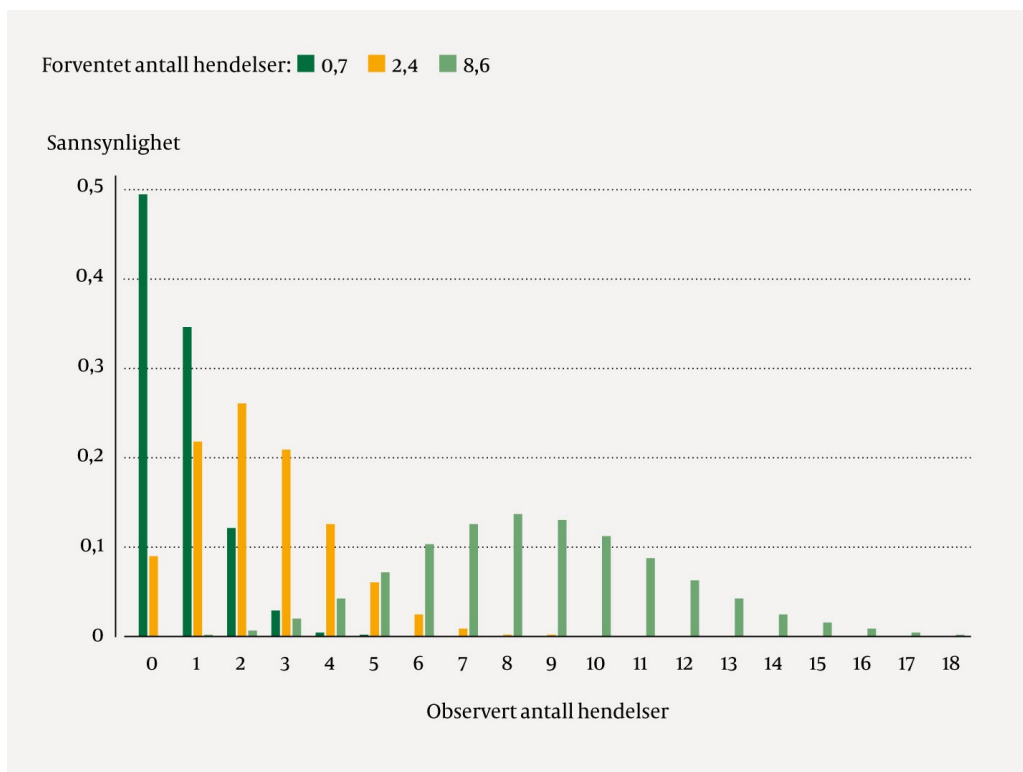
STIAN LYDERSEN

stian.lydersen@ntnu.no

Stian Lydersen er dr.ing. og professor i medisinsk statistikk ved Institutt for psykisk helse, NTNU. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

Når vi vil studere forekomsten av bestemte hendelser i visse tidsrom eller områder, for eksempel fødsler, sykdomstilfeller eller dødsfall i en befolkning, vil poissonfordelingen ofte være egnet.

La oss starte med et eksempel: Barra og medarbeidere studerte fordelingen av antall fødsler med spontan fødselsstart per dag ved Akershus universitetssykehus i perioden 1999–2014 (1). Når hendelser opptrer over tid, slik som her, kan følgende tre antakelser være realistiske: i) Forventet antall hendelser per tidsintervall er konstant, ii) antall hendelser i forskjellige tidsintervall er uavhengige av hverandre, og iii) sannsynligheten for at to hendelser inntreffer nøyaktig samtidig, er forsvinnende liten. Når disse antakelsene gjelder, vil sannsynlighetsfordelingen for antall hendelser i et intervall ha en bestemt form som kalles poissonfordeling (2) (figur 1), oppkalt etter den franske matematikeren Siméon D. Poisson (1781–1840). Når det gjelder fødsler med spontan fødselsstart, synes de tre nevnte antakelsene å være realistiske (1).



Figur 1 Sannsynlighet for observert antall hendelser i forskjellige poissonfordelinger med tre forskjellige verdier av forventet antall hendelser. Gjennomsnittlig forventet antall fødsler per dag var 8,6 i studien til Barra og medarbeidere (1).

Regresjonsanalyse

Ofte vil forventet antall hendelser kunne avhenge av en eller flere variabler. Da kan man benytte regresjonsanalyse i form av poissonregresjon. I den nevnte studien brukte man poissonregresjon til å undersøke om forventet antall fødsler per dag var avhengig av årstall, måned eller ukedag. Det viste seg å være en økende tendens fra 2005 til 2012. Videre var det flest fødsler i juni og juli og færrest fra oktober til januar. Man fant en noe uventet ukedagsvariasjon med færrest fødsler med spontan start på søndager og flest på tirsdager og fredager.

Eksponering

Generelt vil forventet antall hendelser være avhengig av antall personer som kan rammes av hendelsen, eller av over hvor lang tid man teller hendelser. Dette kalles eksponering, og dersom den varierer, må eksponeringen håndteres på en spesiell måte i poissonregresjon. Dette kan illustreres med følgende eksempel: Sommeren 2018 var uvanlig varm i Sørøst-Norge. Ranhoff og medarbeidere undersøkte om det var flere eldre som døde sommeren 2018 sammenliknet med trenden for de ti foregående somrene (3). De studerte dødelighet i de tre sommermånedene i de da 18 fylkene i Norge. Forfatterne brukte en poissonregresjonsmodell med årstall som uavhengig variabel for å ta hensyn til den underliggende trenden i dødelighet over de ti årene, og med en egen uavhengig variabel som var indikator for sommeren 2018. Forventet antall dødsfall i et fylke i en måned vil være proporsjonalt med befolkningsstørrelsen, som i denne studien var antall innbyggere over 75 år. Dette inngikk i analysen som eksponering. Sett under ett fant man ingen overdødelighet i de aktuelle fylkene i Sørøst-Norge i 2018.

Overdispersjon

De fleste sannsynlighetsfordelinger er definert ved to parametere, som for eksempel forventningsverdi (gjennomsnitt) og varians i en normalfordeling. Poissonfordelingen har bare én parameter, som er forventet antall hendelser, og varians er også lik dette. Dette viste seg å stemme meget bra i den nevnte studien av antall fødsler per dag (1, figur 1). I andre situasjoner kan variansen være større enn poissonfordelingen tilsier. Dette kalles overdispersjon og kan i en regresjonsmodell håndteres ved å bruke en såkalt robust varians estimator eller en generalisering av poissonfordelingen som kalles negativ binomisk fordeling (4, s. 553).

Sjeldne hendelser

Poissonfordelingen er spesielt egnet til å modellere sjeldne hendelser, inkludert situasjoner der forventet antall er under 1. Jo høyere forventet antall hendelser er, jo mer vil fordelingen likne på normalfordelingen, slik vi ser i figur 1. Kan man like gjerne bruke normalfordelingen og lineær regresjon når forventet antall hendelser er høyt? Ikke nødvendigvis: Normalfordelingen tar i utgangspunktet ikke høyde for at variansen øker med økende forventningsverdi. Og selv om normalfordelingen kan passe bra rundt «sentrum» av fordelingen, kan den passe dårlig i «utkantene», for eksempel ved dager med høyt antall fødsler, noe man vil ta hensyn til i planleggingen av en fødeavdeling.

REFERENCES

1. Barra M, Lindstrøm JC, Adams SS et al. Sesongjusterte fødselsfrekvenser er Poisson-fordelte. Tidsskr Nor Legeforen 2015; 135: 2154–8. [PubMed][CrossRef]
2. Aalen O, Frigressi A, Moger TA et al. Statistiske metoder i medisin og helsefag. 2. utg. Oslo: Gyldendal akademisk, 2018.
3. Ranhoff AH, Hygen HO, Di Ruscio F et al. Varm sommer 2018 – økt dødelighet blant eldre? Tidsskr Nor Legeforen 2019; 139: 927–30. [PubMed][CrossRef]
4. Agresti A. Categorical data analysis. 3. utg. Hoboken, NJ: Wiley, 2013.

Publisert: 5. januar 2023. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.22.0701

© Tidsskrift for Den norske legeforening 2023. Lastet ned fra tidsskriftet.no 28. januar 2023.