
Romlige data i medisin og helsefag

MEDISIN OG TALL

GEIR AAMODT

geir.aamodt@nmbu.no

Geir Aamodt er professor i epidemiologi ved Institutt for folkehelsevitenskap, Norges miljø- og biovitenskapelige universitet. Hans interesser er romlig epidemiologi og geografiske informasjonssystemer.

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

Vi registrerer ofte geografiske eller stedfestede opplysninger i medisinske og helsefaglige prosjekter. Men hvordan tar vi hensyn til den romlige komponenten når vi skal analysere slike data?

Eksempler på geografiske opplysninger er fylkeshelseundersøkelsene, som blant annet inneholder opplysninger om hvilken kommune, bydel eller postnummer deltakerne bor i. Kliniske studier kan inneholde opplysninger om hvilket sykehus eller hvilken helseregion pasientene er rekruttert fra. Vi er alle knyttet til et geografisk område, og i folkehelsevitenskap vet vi at egenskaper knyttet til dette området har betydning for helsa og trivselen vår. Nærområdet vårt, altså de umiddelbart nærmeste omgivelsene til bostedet, kjennetegnes av ytre fysiske omgivelser, slik som veier, grøntarealer og bebygget område. Men det inkluderer også kulturelle verdier, som deltagelse og sosial kapital i form av nettverk og relasjoner. Nærområdet er derfor en viktig helsedeterminant fordi støy, luftkvalitet, tilgang til parker og grøntområder, kriminalitet og opplevd trygghet påvirker helsa vår på ulike måter. Vi vet også at kvalitetene til nærmiljøet, som tilgang til grønne omgivelser, varierer avhengig av innbyggernes inntekt, utdanning og andre sosiodemografiske faktorer (1).

Romlig avhengighet

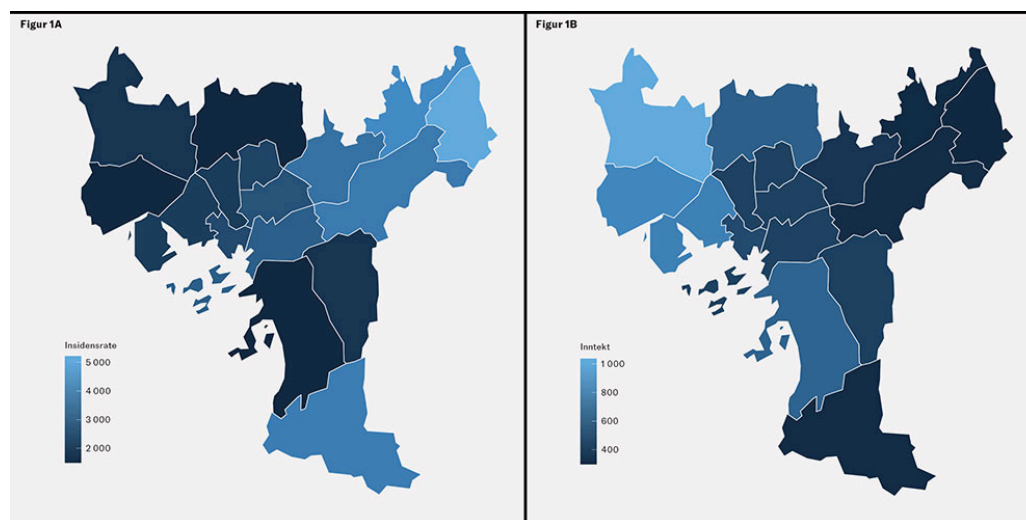
Disse forholdene har betydning for hvordan vi analyserer data der *område* er én av flere opplysninger. Det som påvirker hvordan data skal analyseres, er at områdene er avhengige av hverandre, selv om deltakerne i undersøkelsen er trukket tilfeldig og er uavhengige av hverandre. Egenskaper til et område vil være mer avhengig av

egenskapene til dets naboer enn områder lenger unna. Nærrområder i én kommune har også felles økonomiske betingelser, noe som skiller dem fra nærrområder i nabokommunene.

For å analysere slike data har vi to ulike tilnærminger, avhengig av om vi sammenstiller data for områder eller om data er på individnivå. Vi vil her se på områdedata.

En utvidet lineær regresjonsmodell

Vi starter gjerne en analyse av områdedata med å lage et tematisk kart av utfallsvariabelen, for eksempel forekomsten av sykdom. Figur 1a viser et slikt kart for insidensrate av covid-19 for de første 250 dagene av pandemien i Oslo (2). Vi ser at bydelene Alna og Vestre Aker skiller seg ut som bydelene med høyest og lavest smitte.



Figur 1 Figuren viser insidensrate per 100 000 for covid-19 de 250 første dagene av pandemien a), og gjennomsnittlig inntekt b) for Oslos bydeler. Tallene er hentet fra Folkehelseinstituttet og Statistisk sentralbyrå (2, 4).

Under pandemien var det viktig å få kunnskap om hvordan sosioøkonomiske forhold i de ulike bydelene var knyttet til forekomsten av sykdom (3). Figur 1b viser et kart over gjennomsnittsinntekten for husholdningene i Oslos bydeler. Her ser vi at bydelene Nordre Aker og Alna ligger henholdsvis øverst og nederst på inntektslisten (4).

Det finnes flere modeller for hvordan man kan undersøke slike potensielle sammenhenger, f.eks. modellen som heter *spatial lag model* (5). Det er en utvidelse av en lineær regresjonsmodell. Vi kjenner igjen termene fra en regresjonsmodell, men vi inkluderer et ekstra ledd på høyre side av likningen: $y_i = bx_i + \rho(Wy)_i + e_i$.

Her er utfallsvariabelen vår y_i , mens forklaringsvariabelen er x_i . Parameteren b er regresjonsparameteren vi er interessert i, og den beskriver sammenhengen mellom utfalls- og forklaringsvariablene, mens W er en matrise som holder orden på naboskapet til de ulike bydelene. Matrisen vil ha en positiv verdi for de bydelene som er naboer, og er ellers 0. Parameteren ρ sier noe om styrken i den romlige avhengigheten. Til slutt har vi et feilledd: e_i .

Det som skiller denne modellen fra en lineær regresjonsmodell, er leddet W_{ij} . Dette leddet inneholder opplysninger om utfallsvariabelen fra naboområdene, og det er derfor den har fått navnet «lag». Når vi inkluderer dette leddet, tar vi hensyn til den romlige avhengigheten mellom områdene.

Vi tilpasser nå modellen over med verdiene vist i figur 1, og her ble regresjonsparameteren b lik -3,92 (95 % konfidensintervall -6,11—-1,72). Når gjennomsnittsinntekten økte med 1 000 kroner, ble forventet insidensrate redusert med 3,92 personer per 100 000. Vi tilpasser også en ordinær lineær regresjonsmodell, og her ble resultatet -3,80 (-6,01—-1,54). Regresjonsparametrene i dette eksempelet ble relativt like, og den romlige avhengigheten er derfor også liten. I andre tilfeller kan romlig avhengighet ha større betydning.

Modellen som er vist her, kan tilpasses i programmer som R og Stata. Selv om modellene er mer krevende enn en lineær regresjonsmodell, vil de gi mer presise resultater fordi de tar bedre hensyn til hvordan data genereres.

REFERENCES

1. Aamodt G, Nordh H, Nordbø ECA. Relationships between socio-demographic / socio-economic characteristics and neighborhood green space in four Nordic municipalities – results from NORDGREEN. Urban For Urban Green 2023; 82: 127894. [CrossRef]
2. Folkehelseinstituttet. Surveillance archive covid19. https://github.com/folkehelseinstituttet/surveillance_data Lest 11.11.2024.
3. Nazia N, Law J, Butt ZA. Spatiotemporal clusters and the socioeconomic determinants of COVID-19 in Toronto neighbourhoods, Canada. Spat Spatio-Temporal Epidemiol 2022; 43: 100534. [PubMed][CrossRef]
4. Oslo kommune. Statistikkbanken. <https://statistikkbanken.oslo.kommune.no/> Lest 4.11.2024.
5. Chi G, Zhu J. Spatial regression models for demographic analysis. Popul Res Policy Rev 2008; 27: 17–42. [CrossRef]

Publisert: 13. januar 2025. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.24.0581

Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 14. juli 2026.