

# Populasjon og utvalg i statistikk

---

## MEDISIN OG TALL

ARE HUGO PRIPP

E-post: [apripp@ous-hf.no](mailto:apripp@ous-hf.no)

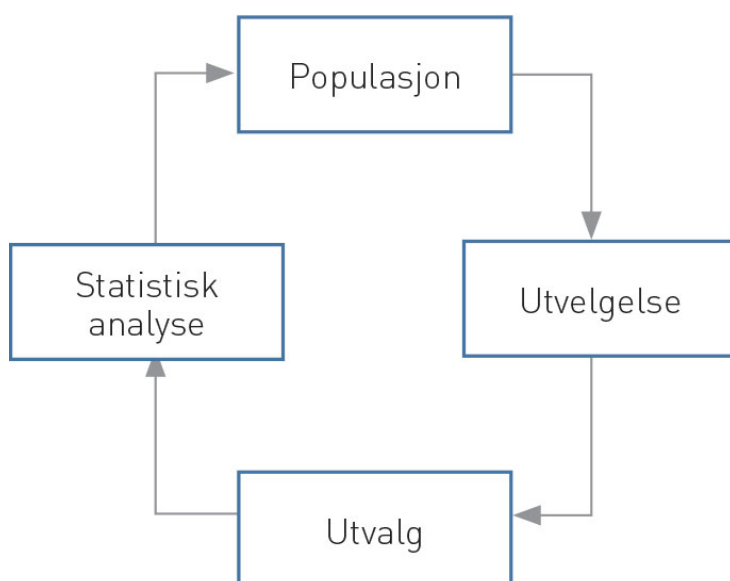
Are Hugo Pripp (f. 1971) er forsker og biostatistiker ved Oslo Centre of Biostatistics and Epidemiology, Forskningsstøtteavdelingen, Oslo universitetssykehus.

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

---

Fordi vi sjelden kjenner egenskapene til en gitt populasjon må vi gjøre antagelser om populasjonen, observere data fra utvalg og anvende statistiske metoder til å trekke slutninger om populasjonen. Analyser av utvalg som gir oss kjennskap til populasjonen, er spesielt betydningsfulle.

I kvantitativ forskningsmetodikk og statistikk tenker vi rent filosofisk at det finnes et endelig sant svar på vårt forskningsspørsmål (1). I statistisk terminologi er dette endelige sanne svaret populasjonens egenskap. Siden vi ikke kjenner denne sannheten, og derfor ikke er «statistisk allvitende», er statistisk analyse av data fra et utvalg det beste vi kan gjøre for å besvare et forskningsspørsmål (fig 1).



**Figur 1** Hensikten med de fleste statistiske analyser er å få informasjon om en populasjon. Ved å velge ut en del av populasjonen får vi et utvalg. Dataene i utvalget analyseres med statistiske metoder. Resultatene fra de statistiske analysene gir informasjon om populasjonens egenskaper

# Populasjon

I dagligtale er en populasjon typisk en gruppe mennesker karakterisert ved for eksempel, kjønn, alder, utdanning, bosted og lignende. I statistikkfaget derimot har begrepet populasjon en mer omfattende betydning. Forenklet sagt er det alle mulige observasjoner, individer eller hendelser vi ønsker å vite noe om. En statistisk populasjon kan være et definert antall, men i statistikkfaget tenker man ofte på det som et hypotetisk uendelig sett med objekter (2).

Det vi ikke vet, må vi anta, noe som er et viktig prinsipp i statistikken. Vi antar vanligvis at populasjonen kan beskrives med såkalte parametere. Disse parameterne angir egenskapene til en matematisk sannsynlighetsfunksjon. Den meste kjente er normalfordelingen. Hvis vi sier at en klinisk variabel er normalfordelt, tenker statistikerne at to parametere som uttrykker henholdsvis forventningen og spredningen, beskriver denne variabelen i populasjonen. Disse parameterne er ofte angitt med de greske bokstavene  $\mu$  og  $\sigma$ .

## Utvalg

Hvis vi visste de nøyaktige verdiene til forventningen og spredningen for en normalfordelt klinisk variabel i populasjonen, var «sannheten» funnet. Hvis vi derimot ikke kan si sikkert at «sannheten» er funnet, er det fornuftig å observere og analysere data fra en del av populasjonen. Denne delen av populasjonen er utvalget. Konseptet med at et utvalg kan gi oss informasjon om hele populasjon kan blant annet tilskrives den norske statistikeren og første direktøren i Statistisk sentralbyrå, Anders Nicolai Kjær (1938-1919) (3).

Et utvalg kan være allerede registrerte data. Hvis såkalte registerdata er utvalget, må vi vurdere kritisk om de gir oss valid informasjon om populasjonen vi ønsker å undersøke. Et eksempel er i hvilken grad utvalget av personer i Helseundersøkelsen i Nord-Trøndelag kan gi oss informasjon om populasjoner andre steder. Hvis det ikke finnes allerede registrerte data, må vi observere data fra et nytt utvalg. Da er det viktig å vurdere sannsynligheten for at en observasjon i populasjonen kan bli del av utvalget. Et tilfeldig utvalg betyr at alle i populasjonen har lik sannsynlighet for å bli del av utvalget. Et utvalg kan også være valgt fordi det er lett tilgjengelig. Et typisk eksempel på dette er sykehusbaserte kontrollgrupper. Da er det spesielt viktig å vurdere i hvilken grad det er et tilfeldig utvalg fra populasjonen vi vil undersøke.

Hvis vi skal undersøke effekten av en ny behandling, er det vanlig å ha et forsøksopplegg. Dataene fra forsøksopplegg er observert under kontrollerte forhold, men i hvilken grad kan de generaliseres til en større populasjon? Deltagere i en randomisert legemiddelstudie kan være annerledes enn alle som kommer til å bruke behandlingen (4). Utvalgets egenskaper er meget viktig informasjon, derfor er ofte den første tabellen i en forskningsartikkel beskrivende statistikk av utvalget.

I prinsippet vil et større utvalg inneholde en større del av populasjonen og dermed være mer representativt. Likevel - hvis det er en skjevhet i utvelgelsen, vil ikke denne skjevheten forsvinne ved kun å inkludere flere observasjoner i utvalget. Dette gjelder spesielt hvis populasjonen ikke er en definert avgrenset gruppe med individer.

## Estimering

For at et utvalg skal kunne gi oss informasjon om parametere til populasjonen, trenger vi estimatorer. En estimator er en statistisk/matematisk funksjon av dataene i utvalget som estimerer parameterverdien i populasjonen. Gjennomsnittet til en kontinuerlig variabel i utvalget estimerer for eksempel forventningen i populasjonen.

Vi inndeler estimeringsmetoder i henholdsvis punkt- og intervallestimeringer. En punktestimering, for eksempel gjennomsnittet, estimerer en unik verdi til parameteren i populasjonen. Intervallestimering, for eksempel konfidensintervallet, estimerer et intervall av mulige (eller sannsynlige) verdier til parameteren.

## Hvordan blir man «statistisk allvitende»?

Det er ikke uvanlig å tro man er blitt «allvitende» når man tolker statistiske analyser fra et godt planlagt studieutvalg med nøyaktige observerte data. En effektiv behandling mot for stor tro på «allvitenhet» er å analysere data fra et nytt utvalg. Ikke sjelden blir konklusjonen om populasjonen litt annerledes enn forrige gang. På det viset fortsetter vi vår møysommelige vandring mot «statistisk allvitenhet».

---

### LITTERATUR:

1. Pripp AH. Vi trenger likestilte forskningsmetoder. *Sykepleien* 2016; 104: 56 - 8.
2. Campbell MJ, Swinscow TDV. *Populations and samples. Statistics at square one.* 11. utgave. London: BMJ Books, 2009: 39-48.
3. Kiær AN. *Den repræsentative undersøgelsesmethode.* Oslo: Statistisk sentralbyrå, 1997.
4. Rothwell PM. External validity of randomised controlled trials: “to whom do the results of this trial apply?”. *Lancet* 2005; 365: 82 - 93. [PubMed][CrossRef]

---

Publisert: 21. august 2017. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.17.0124  
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no