
Nyttige programvarer

MEDISIN OG TALL

ARE HUGO PRIPP

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.

Selv om enkelte statistiske metoder anvendt i medisinsk og klinisk forskning ble utviklet lenge før moderne datamaskiner, er statistikk i dag nær knyttet til statistiske programvarer. Mange viktige fremskritt i statistikkfaget er nå drevet av den teknologiske utviklingen, og den omfattende bruken av statistikk i medisinsk forskning er utenkelig uten datamaskiner og programvarer.

Mekaniske regnemaskiner ble brukt for store datasett så tidlig som i 1930-årene, men de første statistiske programvarene er fra rundt 1970. Mer allmenn bruk av programvarer kom først med de personlige datamaskinene (PC-ene) fra midten av 1980-årene (1). Siden den gang er det utviklet mange programvarer til datahåndtering og statistisk analyse.

I forskning og undervisning ved universitetene, høyskolene og sykehusene er de tre programvarene SPSS, Stata og R mye brukt. De er derfor spesielt relevante for dem som vil arbeide med medisinsk statistikk.

SPSS

Mange forbinder statistisk analyse av medisinske data med SPSS, fordi programvaren er tilgjengelig ved sykehusene og er den mange kjenner fra undervisning og forskning. SPSS er en forkortelse for Statistical Package for the Social Sciences. Den har derfor mange eksempler på bruk av statistikk i samfunnsvitenskap og økonomi, men få eksempler fra medisin og epidemiologi.

De vanligste statistiske metodene finnes i SPSS, men den mangler en del spesialiserte og nyutviklede metoder. Det er heller ikke funksjoner for beregning av statistisk styrke og utvalgstørrelse, noe som er viktig for å planlegge antall deltagere i en studie. En målsetting for SPSS er nok statistisk analyse med minimal eksponering for formler, matematikk og programmering. Jeg har inntrykk av at det i statistiske fagmiljøer med matematisk utspring kan være en viss skepsis til SPSS, men til mange statistiske metoder er den like god og gir de samme resultatene som mer avansert programvare.

Stata

Stata er, slik jeg ser det, en god hybrid mellom en programvare for både avansert matematisk og enkel anvendt bruk av statistikk – den har brukervennligheten til en stor kommersiell programvare og samtidig fleksibiliteten til de mer matematiske og programmeringsbaserte programvarene. Manualer og dokumentasjon er av høy kvalitet, også fra et statistisk vitenskapelig ståsted.

En stor fordel med Stata i forhold til SPSS er muligheten til å laste ned brukerutviklede tilleggsprogrammer. Slike kan være meget nyttig hvis man arbeider med avanserte problemstillinger i klinisk og medisinsk forskning eller trenger andre metoder enn de som følger med fra leverandøren.

R

Det finnes mange gratis programvarer for statistikk. De kan ha varierende brukervennlighet og kan være utviklet for spesifikke fagområder, men en av dem er i ferd med å endre hele virkelighetsbildet når det gjelder statistiske programvarer.

Programvaren R var opprinnelig et programmeringsspråk for statistisk analyse, men har i økende grad overtatt innen både teoretisk og anvendt statistikk. Bruken av programmet er økende, og andre programvarer må nå tilpasse seg R-brukerens behov (2).

R er preget av å være utviklet av og for matematiske statistikere og dataprogrammerere. Programvaren kan derfor være noe overveldende og krevende for dem som kun av og til trenger den til statistiske tester, selv om tilleggsprogrammet Rstudie og brukerutviklede pakker som Rcommander kan forenkle anvendelsen. Min erfaring er også at R fungerer best på datamaskiner med alle rettigheter til å laste ned og installere programvare fra internett.

«Command: Analyse these data and write a manuscript»

Kan det bli utviklet en programvare som automatisk velger passende statistiske analyser for ethvert datasett, presenterer resultatene i ferdige tabeller og figurer samt skriver en tolkning av analysene?

SPSS har allerede noen funksjoner som med varierende hell automatisk velger metode for valgte variabler. For noen tiår siden utførte profesjonelle statistikere de fleste beregninger. Nå kan enhver som vet litt om programvarene utføre statistiske beregninger, selv om korrekt valg av metode og tolkning ikke dermed er garantert. En videreutvikling til en form for automatisk statistisk analyse er ikke utenkelig.

Hvis fremtidige programvarer i større grad kan velge metoder og gi en tolkning av resultater, hvilken rolle får da fremtidens statistikere? Beregninger og tolkninger bør nok fremdeles kvalitetssikres og vurderes. Og høyst sannsynlig blir det økende behov for utvikling av nye statistiske metoder. Når det gjelder medisinsk statistikk og dataanalyse, tror jeg den statistiske og medisinske kompetansen vil bli tettere integrert, heri innbefattet bruk av datamaskiner og programvare.

Det er hevdet at i fremtiden vil diagnostisering, medisinsk behandling og angivelser av prognose være mer basert på analyse av store datamengder og i mindre grad på den enkelte leges kliniske skjønn og erfaring (3). Hvis det går slik, er det viktig at databaser, statistisk analyse, programvare og medisinsk kunnskap stemmer overens. Kanskje det vil føre til at fremtidens statistikere blir litt mer leger og fremtidens leger litt mer statistikere?

LITTERATUR

1. De Leeuw J. Statistical Software - Overview. Technical report, UCLA Department of Statistics, 2009. http://gifi.stat.ucla.edu/janspubs/2009/reports/deleeuw_R_09a.pdf (17.3.2017).
2. Fox J, Leverage A. R and The Journal of Statistical Software. J Stat Softw 2016; 73: 13.
3. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013; 309: 1351 - 2. [PubMed][CrossRef]

Publisert: 30. oktober 2017. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.17.0207

Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 9. juli 2026.