
Kan kunstig intelligens være din statistiske assistent?

MEDISIN OG TALL

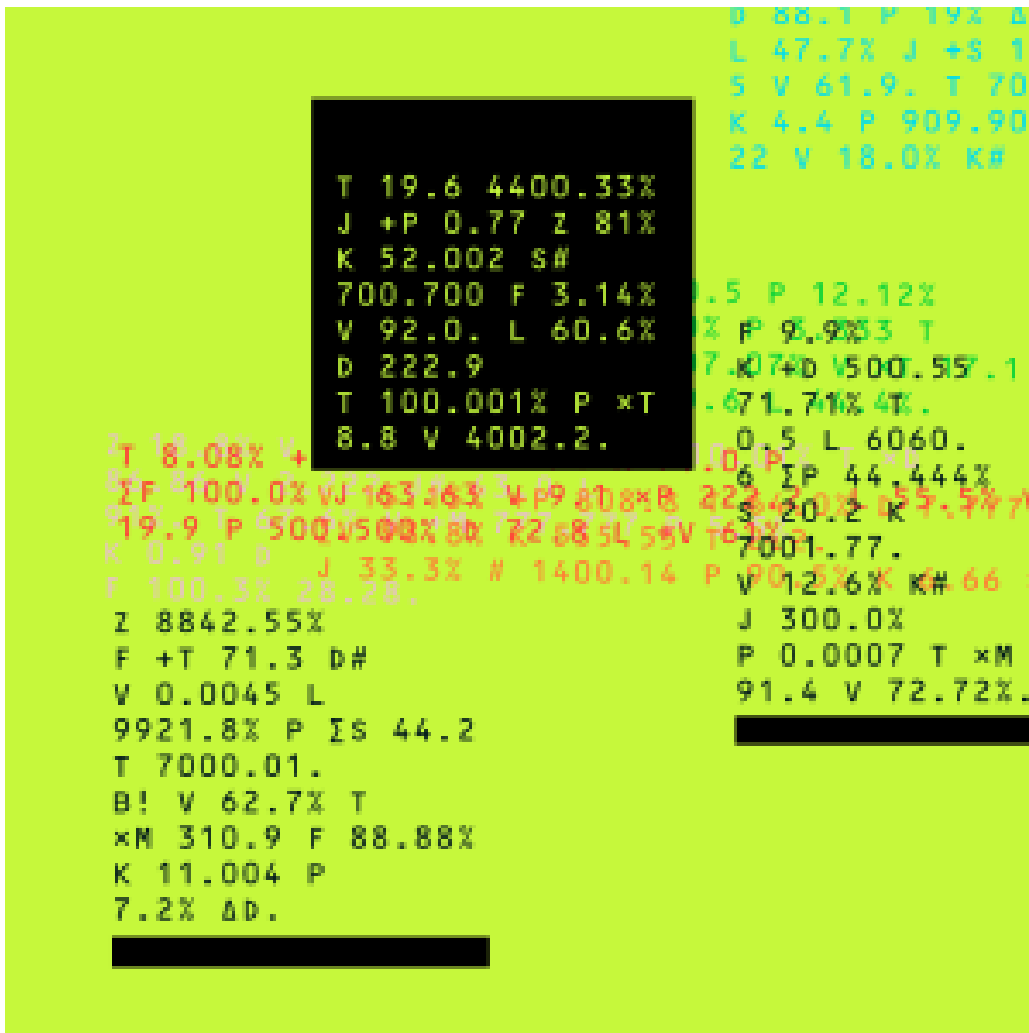
MARIANNE RIKSHEIM STAVSETH

m.r.stavseth@medisin.uio.no

Marianne Riksheim Stavseth er forsker i medisinsk statistikk ved Seksjon for klinisk rus- og avhengighetsforskning (RusForsk) ved Oslo universitetssykehus.

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

Kunstig intelligens (KI) preger stadig mer av medisinenes verden, til både glede og frustrasjon. Men hva skjer når vi slipper teknologien løs på statistikken? Språkmodeller kan skrive kode og forklare analyser på sekunder, men er de til å stole på?



Illustrasjon: Tidsskriftet

Du sitter med et datasett, men husker ikke syntaksen for en blandet modell i Stata eller hvordan du justerer for multippel testing i R. Tradisjonelt sett har løsningen vært tunge manualer eller timevis på Google.

Nå har vi fått store språkmodeller som ChatGPT, Claude og Gemini. Forenklet sagt fungerer disse ved å analysere enorme mengder tekst for å lære mønstre. Når du stiller et spørsmål, søker de ikke i en database etter en fasit. De beregner sannsynligheten for hvilket ord som bør komme etter det forrige. Selv om vi kaller det «kunstig intelligens», *forstår* ikke modellene innholdet. De er eksperter på setningsoppbygging og grammatikk og serverer overbevisende løsninger, men uten reell forståelse. Så før vi overlater p -verdiene helt og holdent til maskinene, kan vi gjøre klokt i å se på hva forskningen faktisk sier om kvaliteten de leverer.

Når språkmodellen velger metode

Et tilbakevendende problem i medisinsk statistikk er studiedesign og valg av metode. I en nylig studie [\(1\)](#) ble det undersøkt om språkmodeller valgte riktig statistisk test i ulike scenarioer. Modellene treffer ofte på enkle spørsmål, men kan ikke sjekke viktige nyanser, for eksempel om de underliggende dataene krever en parametriske eller ikke-parametriske test.

Dette er skummelt fordi modellen svarer med høy selvtillit. En språkmodell som velger feil test, gir deg bare et raskere svar på feil spørsmål. I tillegg har modellene en tendens til å «jatte med» [\(2\)](#). Stiller du et ledende spørsmål som «Er det ikke greit å bruke en t-test her?», vil den ofte bekrefte valget ditt for å være hjelpsom, selv om forutsetningene for testen ikke er oppfylt.

Koden kjører ikke alltid

Hva så med selve kodingen? Selv om svaret fra språkmodellen virker selvsikkert, viser tekniske gjennomganger at den ofte snubler. En artikkel i *Journal of Medical Internet Research* [\(3\)](#) viser at selv om språkmodellene har et stort potensial for dataanalyse, kreves det menneskelig validering for å sikre nøyaktighet. Modellene produserer tidvis kode som enten ikke kjører, eller som gir feil output. Din evne til å kontrollere arbeidet er derfor nøkkelen. La gjerne KI skrive koden, men kun hvis du selv kan validere resultatet.

Et eksempel på trygg bruk er koding av figurer (plot) hvor du allerede kjenner resultatene. Å lage gode illustrasjoner og tabeller kan være tidkrevende. Her kan KI spare deg for mye tid, samtidig som det er lett for deg å vurdere om resultatet ble riktig.

Vær likevel klar over at modellen i verste fall kan hallusinere. Du kan be om en spesifikk R-pakke og få navnet på et bibliotek som ser logisk ut, men som ikke finnes [\(2\)](#). Da bruker du mer tid på å feilsøke modellens fantasi enn på å gjøre jobben selv.

Superkrefter i datavask

Er det da ingenting vi kan bruke dem til? Jo, absolutt. Der KI virkelig glitrer er i arbeidet med «rotete» data [\(4\)](#). Her er nøkkelordet *Regular Expressions* (ofte forkortet Regex). Dette er et kodespråk for å finne mønstre i tekst, men syntaksen er så kryptisk at få forskere behersker den manuelt.

Se for deg et fritekstfelt med blodtrykksmålinger. Én lege har skrevet «120/80», en annen «BP 120 - 80». Å rydde dette manuelt tar dager. En språkmodell kan løse det på sekunder: «Lag et regulært uttrykk i Stata som henter ut tallet før og etter skråstreken, og lagre dem som systolisk og diastolisk trykk.»

Den tålmodige læreren

I tillegg til datavask viser en fersk studie [\(5\)](#) at språkmodeller har et stort potensial som læringsstøtte. For mange er statistikk som et fremmedspråk. Her kan KI fungere som en privatlærer som aldri går lei. Du kan lime inn en

kodesnutt du ikke forstår og be om forklaring linje for linje, eller be om en intuitiv gjennomgang av faste og tilfeldige effekter. Slik endres arbeidsflyten fra blind kopiering til faktisk forståelse.

Fra skriving til verifisering

Kunstig intelligens kan være en fantastisk statistisk assistent. Den er utrettelig, kan alle programmeringsspråk og er alltid tilgjengelig. Men den kan finne på å lyve, bruke feil metode og dikte opp referanser. Rollen din som forsker endres derfor fundamentalt. Du går fra å være den som velger testen eller skriver koden, til å bli den som må verifisere den (6). Du må lese koden kritisk, sjekke forutsetningene og tolke resultatene. Sist, men kanskje viktigst av alt: aldri dele pasientsensitive data med en språkmodell!

I arbeidet med å skrive denne teksten har forfatteren brukt språkmodellen Gemini som en diskusjonspartner for strukturering, idémyldring og språkvask. Forfatteren tar det fulle ansvaret for at innholdet i teksten er korrekt.

LITTERATUR

1. Mondal H, Mondal S, Mittal P. Evaluating large language models for selection of statistical test for research: A pilot study. *Perspect Clin Res* 2024; 15: 178–82. [PubMed][CrossRef]
2. Thirunavukarasu AJ, Ting DSJ, Elangovan K et al. Large language models in medicine. *Nat Med* 2023; 29: 1930–40. [PubMed][CrossRef]
3. Ruta MR, Gaidici T, Irwin C et al. ChatGPT for Univariate Statistics: Validation of AI-Assisted Data Analysis in Healthcare Research. *J Med Internet Res* 2025; 27: e63550. [CrossRef]
4. Stavseth MR. Husk å vaske dataene før bruk! *Tidsskr Nor Legeforen* 2023; 143. doi: 10.4045/tidsskr.22.0825. [PubMed][CrossRef]
5. Roselló Atanet I, Tomova M, Sieg M et al. Generating learning guides for medical education with LLMs and statistical analysis of test results. *BMC Med Educ* 2025; 25: 458. [PubMed][CrossRef]
6. Lin A, Wang Z, Jiang A et al. Large language models in clinical trials: applications, technical advances, and future directions. *BMC Med* 2025; 23: 563. [PubMed][CrossRef]

Publisert: 10. mars 2026. *Tidsskr Nor Legeforen*. DOI: 10.4045/tidsskr.26.0092
Opphavsrett: © Tidsskriftet 2026 Lastet ned fra tidsskriftet.no 23. juni 2026.